平成 28 年度

最新の IT 技術を活用した
デジタルアーカイブ・システムの調査検討

報　告　書

平成 29 年 1 月

インフォコム株式会社

内容

1. 調査検討の概要 ..... - 1 -

1.1. 背景と目的 ..... - 1 -

1.2. 調査計画 ..... - 1 -

1.2.1. 調査検討の枠組み ..... - 1 -

1.2.2. 調査項目の設定 ..... - 2 -

1.2.3. 調査検討委員会 ..... - 3 -

1.3. 調査検討要旨 ..... - 4 -

1.3.1. デジタルアーカイブ・システムに関連する最新 IT 技術に係る調査検討 ..... - 4 -

1.3.2. デジタルアーカイブ・システムで提供するサービスに係る調査検討 ..... - 7 -

2. デジタルアーカイブ・システムに関連する最新 IT 技術に係る調査検討 ..... - 9 -

2.1. システム構成等に係る調査 ..... - 9 -

2.1.1. システム構成要素 ..... - 9 -

2.1.2. 運用形態 ..... - 15 -

2.1.3. デジタルアーカイブ・システムへの適用 ..... - 22 -

2.2. クラウド等 IT 技術に係る調査 ..... - 28 -

2.2.1. クラウド技術 ..... - 28 -

2.2.2. サーバ環境 ..... - 31 -

2.2.3. ネットワーク環境 ..... - 35 -

2.2.4. バックアップ関連技術 ..... - 38 -

2.2.5. ネットワークセキュリティ技術 ..... - 43 -

2.2.6. 目録関連技術 ..... - 46 -

2.2.7. 目録データの提供 ..... - 54 -

2.2.8. 情報検索技術 ..... - 58 -

2.2.9. デジタルコンテンツ配信技術 ..... - 64 -

2.2.10. クライアント環境 ..... - 68 -

2.3. デジタルアーカイブ・システムに活用可能と考えうる技術に係る調査 ..... - 74 -

2.3.1. OCR 技術と画像検索技術 ..... - 74 -

2.3.2. 自動翻訳 ..... - 77 -

2.3.3. 辞書/シソーラス検索機能と利用可能な辞書データ ..... - 78 -

2.4. デジタルアーカイブ・システムへの適用に係る検討 ..... - 79 -

2.4.1. システム構成等のデジタルアーカイブ・システムへの適用 ..... - 79 -

2.4.2. クラウド等 IT 技術のデジタルアーカイブ・システムへの適用 ..... - 81 -

2.4.3. デジタルアーカイブ・システムに活用可能と考えうる技術 ..... - 87 -

3. デジタルアーカイブ・システムで提供するサービスに係る調査検討 ..... - 88 -

3.1. デジタルアーカイブ・システムで提供するサービスに係る調査 ..... - 88 -

3.1.1. 概要 ..... - 88 -

3.1.2. 全国の公文書館等へのアンケート ..... - 89 -
3.1.3. デジタルアーカイブの基本機能(利用者向け) ..... - 108 -
3.1.4. デジタルアーカイブの基本機能(業務) ..... - 137 -
3.1.5. 利用者拡大のための機能・サービス ..... - 140 -
3.1.6. 利用サービス向上のための機能・サービス ..... - 149 -
3.2. デジタルアーカイブ・システムへの適用に係る検討 ..... - 152 -
3.2.1. デジタルアーカイブの基本機能(利用者向け) ..... - 152 -
3.2.2. デジタルアーカイブの基本機能(業務) ..... - 157 -
3.2.3. 利用者拡大のための機能・サービス ..... - 160 -
3.2.4. 利用サービス向上のための機能・サービス ..... - 163 -
4. まとめ ..... - 164 -
4.1. 検討結果 ..... - 164 -
4.1.1. 「デジタルアーカイブ・システムに関連する最新 IT 技術に係る調査検討」の調査検討結果のまとめ. - 164 -
4.1.2. 「デジタルアーカイブ・システムで提供するサービスに係る調査検討」の調査検討結果のまとめ. - 167 -
4.2. 標準仕様書の改訂に向けて ..... - 169 -
4.2.1. 標準仕様書の改訂に向けた考え方 ..... - 169 -
4.2.2. 「標準仕様書の考え方」の改訂案 ..... - 170 -

※ 本報告書は独立行政法人国立公文書館による委託事業としてインフォコム株式会社が実施した「最新の IT 技術を活用したデジタルアーカイブ・システムの調査検討業務」の成果を報告するものである。

最新のIT技術を活用したデジタルアーカイブ・システムの調査検討結果（概要）

【1】デジタルアーカイブ・システムに関連する最新IT技術に係る調査検討

主な検討事項	調査結果（検討の方向性）	検討結果	標準仕様書の改訂に向けた見直しポイント
1.　システム構成	システム基盤：　自館設置，データセンター設置，プライベートクラウド（IaaS，PaaS）の利用 アプリケーション提供方法：　パッケージ導入等，サービス利用型（SaaS）の利用 運用形態：　単館運用，共同運用	クラウドサービスを利用した共同運用の場合、付加価値提供にはプライベートクラウド、負担軽減にはサービス利用型（SaaS）が有効。	各機関の要望に応じたシステム構成や運用形態を選択可能とする。 （従来は機能に応じた3種類の構成を提示。今後は機能とインフラの組み合わせによる構成を提示。）
2.　クラウド等IT技術	（1）情報システムの基盤にかかるIT技術 クラウド技術：　様々なクラウドサービス（IaaS，PaaS，SaaS） セキュリティ：　HTTPSによる暗号化，IPA等公表資料で示される対策	状況に応じた様々なクラウドサービスが活用可能。セキュリティ対策が必要。	クラウド環境を活用する際の留意点、考えられるセキュリティ対策を示す。 （HTTPS、WAF 等）
	（2）アーカイブズの目録や検索にかかるIT技術 メタデータ関連技術： 最新版EAD，Dublin Coreの拡張項目，アプリケーションプロファイルの適用 目録データの提供： パーマリンクによるアクセスの永続性，オープンデータのためのWebAPI 情報検索技術： 全文検索とRDBの併用，横断検索プロトコル（SRU/WやOpenSearch）	相互運用性を高め、二次利用を容易にするため、メタデータやデータ連携等の標準化を推進。	国内外の標準規格等の最新版への準拠やオープンデータへの対応を考慮する。 （EAD3、OpenSearch、WebAPI 等）
		長期安定利用のために永続してアクセスできる仕組みが重要。	目録やデジタルコンテンツへのアクセスの永続性を考慮する。 （パーマリンク（固定URL）の付与、リダイレクト 等）
	（3）Webアプリケーションとしての側面にかかるIT技術 デジタルコンテンツ配信：　マルチメディア配信 クライアント環境：　マルチデバイス対応，モバイル通信	利便性向上のため、マルチメディア、マルチデバイスに対応。	音声や動画の配信を考慮する。 多様な利用者環境に対応する。 （マルチデバイスへの対応 等）
	（4）その他活用可能と考え得る技術 OCRと画像検索（くずし字検索），自動翻訳等	利便性向上に資する研究成果を把握。実用に向けた今後の進展を期待。	技術の進展などにより、導入可能な場合は追加する。 （OCRによる全文テキストデータ化、自動翻訳による複数言語での情報提供 等）

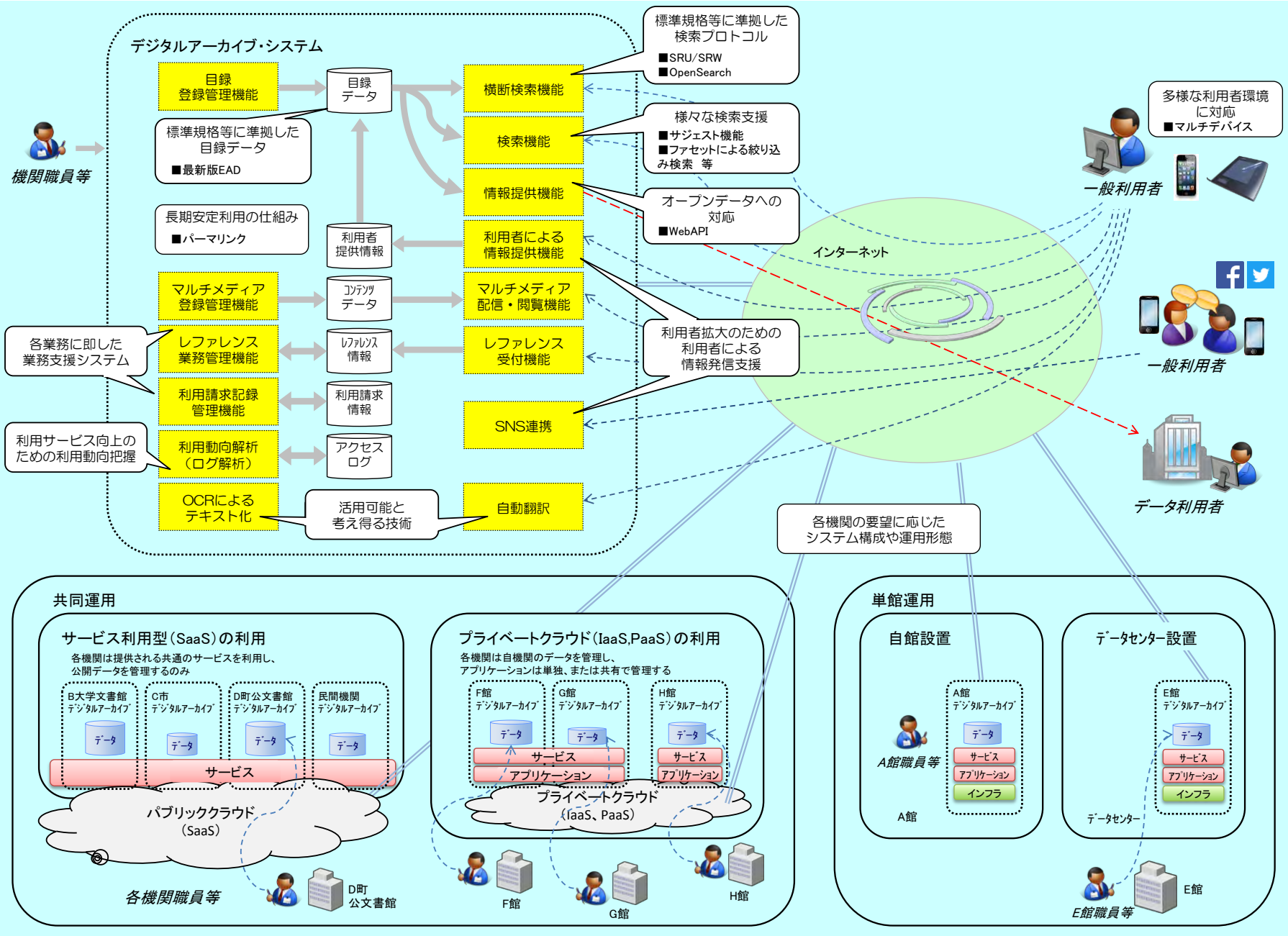
【2】デジタルアーカイブ・システムで提供するサービスに係る調査検討（アンケート調査）

主な検討事項	調査結果（検討の方向性）	検討結果	標準仕様書の改訂に向けた見直しポイント
1.　デジタルアーカイブの基本機能(利用者向け)	検索支援機能（サジェスト機能，異体字等検索機能），動画・音声資料の提供	想定する利用者に合わせた検索支援機能の充実、マルチメディアコンテンツへの対応。	様々な検索支援機能を選択可能とする。 （検索キーワードのサジェスト機能、異体字検索機能 等） 音声や動画の配信を考慮する。
2.　デジタルアーカイブの基本機能(業務)	利用請求記録管理機能，レファレンス業務の管理機能，資料の利用制限機能	各種業務に即した業務システムの整備が求められている。	各種業務をシステム化する際の考え方を示す。 （レファレンス業務に係る情報（受付内容、回答内容等）をデータベース化 等）
3.　システム導入に係る事情	運用形態： 同一組織内共同（39%），他の公文書館と共同（38%），単館運用（18%） 共同運用時のシステム構成：　インフラ・アプリ共有が6割以上 システム導入の制約：　予算，人，システム，制度	機関の要望に応じて適切なシステム構成や運用形態を選定。	各機関の要望に応じたシステム構成や運用形態を選択可能とする。 （インフラ・アプリを共有する構成、単館運用による構成 等）

【3】デジタルアーカイブ・システムで提供するサービスに係る調査検討（文献等調査）

主な検討事項	調査結果（検討の方向性）	検討結果	標準仕様書の改訂に向けた見直しポイント
1.　デジタルアーカイブの基本機能(利用者向け)	検索時の利便性向上：　辞書機能，サジェスト機能，ファセットによる絞り込み検索 ユーザインターフェースの高機能化：　複数資料の重ね合わせ，レスポンシブウェブデザイン	検索支援や利用環境の充実を図る機能の整備。	様々な検索支援機能を選択可能とする。 （検索結果を絞り込むためのキーワード等を示す機能 等） 多様な利用者環境に対応する。 （レスポンシブウェブデザイン 等）
2.　デジタルアーカイブの基本機能(業務)	目録データ登録管理機能，画像データ管理機能，利用状況把握機能，閲覧業務関連機能	各種業務に即した業務システムの整備が求められている。	各種業務をシステム化する際の指針を示す。 （利用状況把握機能、閲覧業務関連機能 等）
3.　利用者拡大のための機能・サービス	SNS連携，利用者からの提供情報（メタデータ，画像データ）を【直接的/間接的に】登録する機能，デジタルコンテンツの展示・広報への活用，他機関との連携	利用者がデータ登録や情報発信に寄与する機能は利用者拡大に有効（運用方針の明確化が必要）。	利用者拡大のための機能を選択的に導入可能とする。 （SNS連携、利用者による資料紹介の支援機能 等）
		資料情報の展示等への二次利用は利用者拡大に有効。	
4.　利用サービス向上のための機能・サービス	レファレンスサービス，利用動向解析機能（システムのログ解析）	レファレンスサービスの実施を支援する機能は利用サービス向上に有効。	利用サービス向上のための機能を選択的に導入可能とする。 （レファレンス情報の公開、システムログの解析による利用動向の把握 等）
		利用動向解析機能は利用サービス向上に有効。	

最新のIT技術を活用したデジタルアーカイブ・システムの調査検討結果のイメージ



1. 調査検討の概要

1.1. 背景と目的

独立行政法人国立公文書館では、平成17年4月より、「国立公文書館デジタルアーカイブ」を運用している。同館では、同システムの運用実績を踏まえ、全国の公文書館等におけるデジタルアーカイブ化の推進に資するため、「全国の公文書館等におけるデジタルアーカイブ・システムの標準仕様書」（平成 21 年 3 月作成、平成24年10月一部改訂）を作成し、全国の公文書館等へ配布するとともに、訪問説明等を実施してきた。

一方、平成 27 年度全国公文書館長会議において、「「所蔵資料等のデジタル化」に取り組む基本的考え方」（平成 27 年 6 月 9 日全国公文書館長会議）が取りまとめられ、クラウド技術等を活用した共同利用型システムによるデジタルアーカイブの構築について積極的に推進する旨が言及された。

また、内閣府に設置された「国立公文書館の機能・施設の在り方等に関する調査検討会議」による「国立公文書館の機能・施設の在り方に関する基本構想」（平成 28 年 3 月 31 日）においても、同館による全国の公文書館等への技術支援と合わせ、デジタルアーカイブ化の推進におけるクラウド技術等の活用についての言及がある。

同館では、このような状況を踏まえ、最新の IT 技術を活用したデジタルアーカイブ・システムに関する情報を提供し、全国の公文書館等のデジタルアーカイブ化の推進に資するため、

- デジタルアーカイブ・システムに関連する最新の IT 技術
- デジタルアーカイブ・システムで提供するサービス

の調査検討を行うこととした。

1.2. 調査計画

1.2.1. 調査検討の枠組み

本調査検討では、最新の IT 技術を活用したデジタルアーカイブ・システムについて、デジタルアーカイブ関連技術の調査と、デジタルアーカイブで提供するサービスの両面から検討するため、以下に挙げる３つのアクションを実施することとする。

(1) デジタルアーカイブ・システムに関連する最新の IT 技術に係る調査検討

デジタルアーカイブ・システムに関連する最新の IT 技術について、動向や事例等を調査し、「活用の可能性」「稼働環境」「求められる技術要件」「コスト」「課題」等を勘案し、デジタルアーカイブ・システムへの適用を検討する。

(2) デジタルアーカイブ・システムで提供するサービスに係る調査検討

デジタルアーカイブ・システムで提供するサービスについて、動向や事例等を調査し、「導入の容易さ」「運用効率」「利活用促進効果」「課題」等を勘案し、デジタルアーカイブ・システムへの適用を検討する。また、全国の公文書館等にアンケートを実施し、求められる機能・サービスについて、意見等を確認す

る。

(3) 調査結果に基づくデジタルアーカイブ・システムへの適用に係る検討

上記(1)、(2)の調査結果を踏まえ、デジタルアーカイブ・システムへの適用を検討し、併せてデジタルアーカイブ・システムの標準仕様書の改訂を見据え、同仕様書改訂版の骨子を案としてとりまとめる。

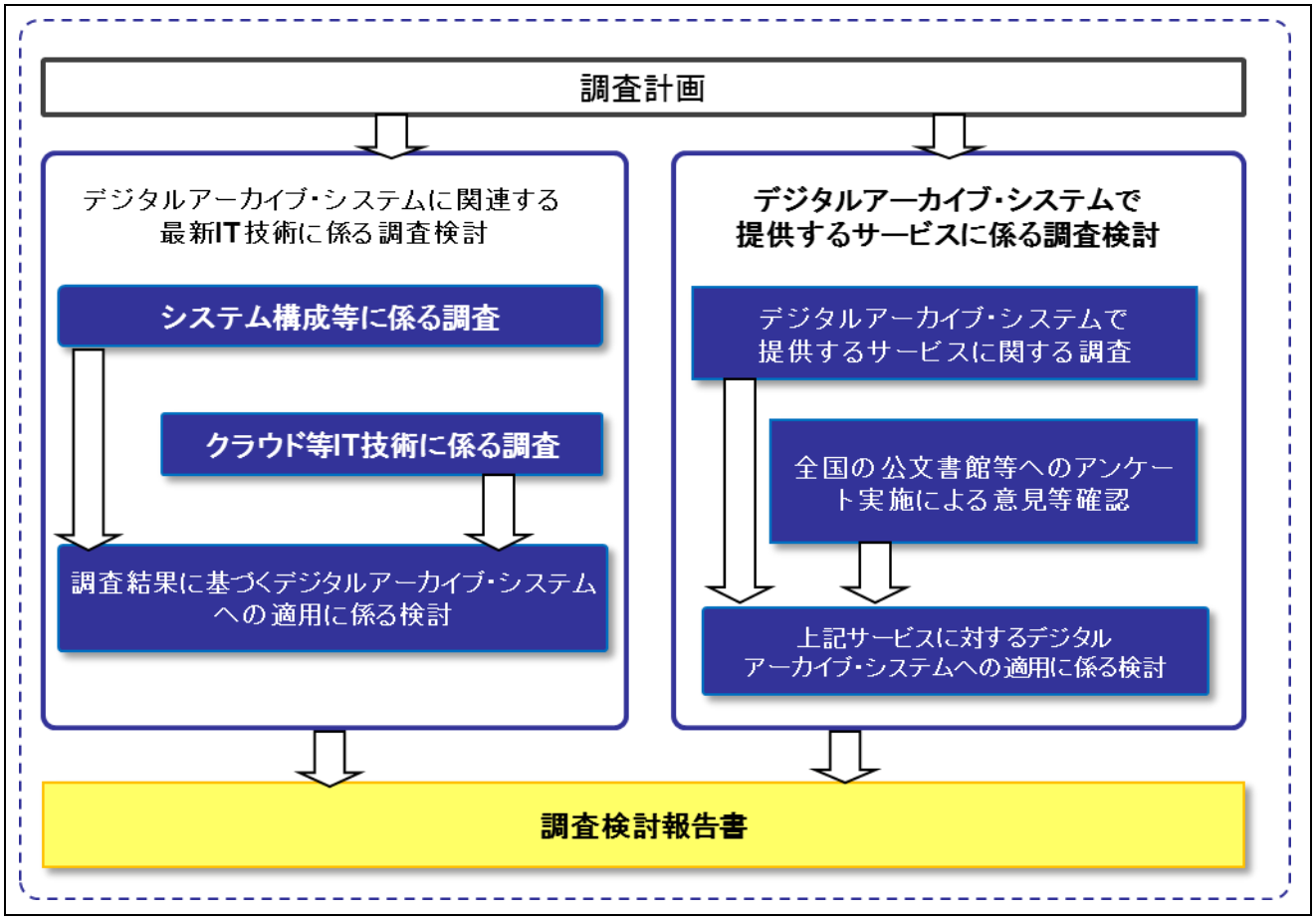


図 1 調査検討全体の流れ

1.2.2. 調査項目の設定

本調査検討において、調査する技術・サービス等は、デジタルアーカイブ・システムに関する技術や考え方が整理されているデジタルアーカイブ・システムの標準仕様書を踏まえて、以下のとおり設定する。

(1) デジタルアーカイブ・システムに関連する最新の IT 技術に係る調査検討

① システム構成等に係る調査

システム基盤、アプリケーションの提供方法、運用形態の観点で、デジタルアーカイブ・システムに適用されるシステム構成を調査する。

② クラウド等 IT 技術に係る調査

デジタルアーカイブ・システムに関する基本的な技術として、クラウド技術、サーバ環境、ネットワーク環境、バックアップ関連技術、ネットワークセキュリティ技術、目録関連技術、目録データの提供、情報検索技術、デジタルコンテンツ配信技術、クライアント環境について、最新の動向等を調査する。また、

デジタルアーカイブ・システムにおいて活用可能と考えうる技術として、OCR 技術と画像検索技術、自動翻訳、辞書/シソーラス検索機能と利用可能な辞書データを調査する。

(2) デジタルアーカイブ・システムで提供するサービスに係る調査

① 全国の公文書館等へのアンケート実施による意見等確認

デジタルアーカイブ・システムに求められる機能として、利用者向けサービスの観点、業務や運用に係る機能の観点の両面から、デジタルアーカイブ・システムに求められる機能・サービスをアンケートにより調査する。

また、望ましいシステムの運用形態として、機関が単独で運用するシステム構成のみならず、共同運用型システムを含めた、望ましいシステムの運用形態をアンケートにより調査する。

② 国内外のデジタルアーカイブ・システムが提供するサービスに係る調査

全国の公文書館等及び民間、海外の代表的なデジタルアーカイブを対象として、デジタルアーカイブ・システムの基本機能(利用者向け)、デジタルアーカイブ・システムの基本機能(業務)、利用者拡大のための機能・サービス、利用サービス向上のための機能・サービスの観点で、動向及び特徴的事例について調査する。

1.2.3. 調査検討委員会

以下の 3 分野に該当する専門家(各分野 1 名、計 3 名)からなる調査検討委員会により、調査検討結果に対する専門的な見地からの助言等を得る。

表 1 調査検討委員会 委員名簿

専門分野	氏名、所属等
インターネット全般に係る IT 技術・サービス動向に関する専門家	国立情報学研究所 准教授 大向 一輝
デジタルアーカイブに係る IT 技術・サービス動向に関する専門家	筑波大学 講師 永森 光晴
デジタルアーカイブの利用・連携に関する専門家	東京大学文書館 准教授 森本 祥子

1.3. 調査検討要旨

1.3.1. デジタルアーカイブ・システムに関連する最新 IT 技術に係る調査検討

デジタルアーカイブ・システムに関連する最新 IT 技術について調査し、調査した技術を項目ごとに分けた上で、それぞれの項目におけるデジタルアーカイブ・システムに対する適用可能性について検討した。調査したIT 技術については、技術面・コスト面においてその多くがデジタルアーカイブ・システムへ適用が可能であると考えられるが、解決すべき課題も多い。以下に項目別の検討状況を示す。

システム構成	デジタルアーカイブ・システムのシステム構成には、一つの機関が単独で構築・運用する形態（単館運用）と複数の機関が共同して構築・運用する形態（共同運用）が考えられる。 単館運用の場合、業務や利用者の要求にあわせたアプリケーションを構築し、サービス面において機関の独自性を出すことが比較的容易であるが、予算や人員を独自に確保する必要がある。また、システムの設置形態として、オンプレミス型、コロケーション型（データセンター活用型）が考えられるが、デジタルアーカイブ・システムが、インターネットを通じて 24 時間サービス提供することを前提とすると、安定した電源や回線等を備え、運用担当者が 24 時間体制で監視をするデータセンターを利用することがより望ましく、コロケーション型がより適していると言える。 共同運用の場合、利用者の利便性向上という点からは、単にシステム基盤が共有されているというだけではなく、アプリケーションにおいても共通した機能の利用が求められるため、アプリケーションの開発・導入やパッケージ導入においては、利便性向上につながる、共同運用ならではの機能を織り込むことや、共同運用する機関の事情に配慮するなど、柔軟な対応を要する。このような付加価値を提供する観点から、システム基盤としてはパブリッククラウドよりも、専用のクラウド環境であるプライベートクラウドの採用が適していると考えられる。一方、費用や人員等の負担軽減という点からは、サービス利用型（SaaS）システムの活用が有効な選択肢である。サービス利用型のデジタルアーカイブ・システムを利用する場合は、プライベートクラウドよりもパブリッククラウドを採用する方が、提供されているサービスの選択肢も多く、コストメリットが現れやすい。
クラウド技術	低コストでシステムを導入・運用できることから有用であり、また、様々なサービスが商用ベースで提供されている。しかし、データのセキュリティや活用可能なアプリケーション機能等の運用環境において、それぞれの機関が提供したいと考えるサービスの実現可能性や運用方法等の十分な検討を要するものと考えられる。
サーバ環境	サーバ機器、OS、ミドルウェアは、低価格化とオープンソース化が進んでいる。また、サーバ機器の性能向上に合わせて仮想化技術も発展している。ストレージの容量あたりの単価は減少傾向にあり、大容量のデータを扱うためのコスト面でのハードルは低くなりつつある。また、クラウドストレージを使うことで、提供

	されるサービス内容に制限はあるものの、必要な容量のみをコストを抑えて用意することが可能である。このような状況から、デジタルアーカイブ・システムを実現するためのサーバ環境は年々高機能かつ低コスト化が進んできていることが伺え、サーバ環境（ハードウェア面）の整備におけるコスト面でのハードルは低減しているものと考えられる。
ネットワーク環境	インターネット接続回線には、公開するコンテンツ、想定のアクセス数、予算に応じた適切な回線を用意することが必要である。ただし、各機関がすでにインターネット回線を使用しており、その回線が相当の要件を満たしている場合は、セキュリティに配慮して、既存の回線の活用を図ることが合理的な対応であると考える。 通信プロトコルにおいては、HTTP の最新版である HTTP/2 への対応により、閲覧速度向上により利用者のストレス軽減という効果が期待できる。また、利用者やサービス提供者（各機関）が送受する情報に係るセキュリティ等を考慮すると、暗号化した通信を提供する HTTPS への対応が望ましいと考えられる。
バックアップ関連技術	現時点において、バックアップ関連技術としては、複数のバックアップ方式、バックアップメディア等が様々な企業より提供されている。各機関の保有するデータ量と増加予測、データの特徴に応じて、適切なバックアップ方式やメディアを検討し導入することが推奨される。
ネットワークセキュリティ技術	コンピュータネットワークにおける脅威は年々増大し、また、多様化の一途を辿っている。このような状況におけるネットワークセキュリティへの対応としては、これらの脅威に対抗する様々な技術やサービスが提供されている。しかし、これらの技術等の導入に係るコストは、セキュリティレベルに応じて高くなる傾向にある。各機関のセキュリティ対策方針等に応じて IPA 等の公表資料等で示される対策を参考に適切な対応を選択する必要がある。
目録関連技術	符号化記録史料記述の国際的な標準の最新版である EAD3 への対応が望ましいが、国際公文書館会議で検討されている新しい記述標準である RiC(Records in Contexts)の動向も注視すべきである。横断検索では Dublin Core を採用することは有用である。また Dublin Core の拡張項目である DC タームスでは、より精度の高い情報の記述が可能である。横断検索を実現する機能の運用や検索精度の面から、最低限共通する目録データ項目を設定した上で、連携する目録データ項目を整理することやアプリケーションプロファイルの考えを取り入れることで相互運用性を高めることが重要である。 また、デジタルアーカイブ・システムにおいて、システム移行等を経ても、目録データ等提供情報を長期的に安定して利用可能とするパーマリンク（固定 URL）の導入は、同システムにおける目録情報等を識別する技術であるとともに、利用者の利便性を損ねないための技術として有用である。
目録データの提供	人間が利用するためのデータ提供と機械可読性の高い形式でのデータ提供の両者に対応することが推奨される。後者においては、デジタルアーカイブ・システムが提供するデータを別のアプリケーションプログラムで取得、活用可能とする

	ことにより、利用者や他機関が作成するアプリケーションから、各機関の目録データやデジタルコンテンツデータを利用することが可能となり、利活用の拡大を期待することができる。更にデジタルアーカイブ・システムで提供するデータは、今後オープンデータへの対応が求められると考えられる。 また、利用者が安心してデジタルアーカイブ・システムが提供するデータを利用できるようにするためには、政府標準利用規約やクリエイティブ・コモンズといったライセンスに関する標準への準拠と提供データにおける表示に対応することが望ましい。
情報検索技術	デジタルアーカイブ・システムで扱うデータは非構造化データ等多様化が進んでおり、このような状況における情報検索の手段として、全文検索への対応がある。全文検索に対応する仕組み（全文検索エンジン）には、高機能・高性能なものがオープンソースでも提供されており、導入のハードルは決して高いものではなくなってきている。また、全文検索機能を有する RDBMS を用いることにより、データ管理の効率化を図る方法も考えられる。横断検索では、SRU/W、OpenSearch のいずれかを実装することが、相互運用性の向上という点からも推奨される。また、オープンデータ推進の観点では、機械可読形式でのデータ提供と併せて WebAPI を提供することが推奨される。
デジタルコンテンツの配信	デジタルアーカイブで扱うコンテンツの多様化を踏まえて、動画・音声等マルチメディアの配信について検討する必要がある。これらの配信方式としては、ストリーミング方式やダウンロード方式がある。動画・音声の配信サービスでは、データ保護等の観点から、ストリーミング方式によることが一般的であるが、配信に専用のソフトウェアを必要とする。また、配信サーバの負荷軽減や利用者の閲覧等におけるレスポンスの向上を期待することができるコンテンツキャッシュ技術の適用が可能である。しかし、配信ソフトウェアやコンテンツキャッシュ技術の導入は、導入コストや運用コストの発生を伴うことから、費用対効果を踏まえた検討が必要であると考える。
クライアント環境	普及の著しいスマートフォン、タブレットといったスマートデバイスによる利用の増加が予想されるため、マルチデバイスに対応したユーザーインターフェースを備えることが望ましい。 また、高速モバイル通信を用いたスマートデバイスの利用が増加しており、クライアント環境として今後も考慮すべきところであるが、通信回線におけるトラフィックが増大していることに対し、月当たりのデータ通信量の上限を設ける等の対応を実施する通信キャリアが出てきている。利用者の利便性を損なわないように配慮した対策等を講じられることが望まれる状況にある。
デジタルアーカイブ・システムに活用可能と考えうる技術	OCR や画像検索、自動翻訳といった技術が研究されており、それぞれ成果を挙げ、実用段階に近づいている。今後の進展が期待されるところである。活用可能と考え得る技術は、技術の進展などにより導入可能かを検討し、適用を判断することになると考える。

1.3.2. デジタルアーカイブ・システムで提供するサービスに係る調査検討

(1) 全国の公文書館等へのアンケート結果

全国の公文書館等へのアンケート結果を分析し、以下の四点に取りまとめた。

① 整備が期待される機能（サービス）
② 考慮すべき対外的な対応
③ 業務支援機能の充実
④ システム導入・運用の低コスト化

取りまとめた内容から、全国の公文書館等、とりわけ市区町村レベルではシステム化の予算と人員の確保が大きな課題となっているということを確認した。また、機関が保有するデータを外部で管理することに制度上の制約があるとした機関がある一方、インフラとアプリケーションを共有してもよいと回答した機関が 6 割以上であり、その課題の解決策として、アプリケーションも含めた安価な共同利用型システムの活用を期待している機関も少なからず確認された。ただし、運用形態で単館運用を希望する機関も一定数あり、機関の要望に応じて適切なシステム構成や運用形態を選択できることが必要であることが確認された。

整備が期待される機能（サービス）では、サジェスト機能や異体字等の検索機能などの検索支援機能や、動画・音声資料の提供機能が挙げられた。

業務に係る機能（サービス）の面では、利用請求記録管理機能やレファレンス業務管理機能、職員を含めた利用者に対する資料の利用制限に対する要望が大きかった。特に共同利用型システムで実現する機能を検討する際にはこれらの機能（サービス）の提供を考慮するとともに、これらの機能（サービス）の実装と運用方法について、システム運用機関における運用の実際や利用者の操作挙動を充分に想定することが重要であると言える。

(2) 国内外のデジタルアーカイブ・システムの調査結果

国内外のデジタルアーカイブ・システムに対し、それぞれのシステムが提供するサービス（機能）について調査し、その状況を以下の五点に取りまとめた。

① 所蔵資料の目録検索サービス提供状況
② 所蔵資料のデジタル画像提供状況
③ 動画・音声を含むデジタルコンテンツ提供状況
④ ソーシャルメディア（SNS 等）の活用事例
⑤ 利用者参加型データベース整備事例

取りまとめた内容から、目録データの検索サービスをインターネット上で提供している機関は調査対象の 5 割弱であり、デジタルアーカイブといえど決して当たり前に提供されているサービスではないこと、一方、5 割を超える機関で所蔵資料のデジタル画像を公開しており、利用者にアピールするコンテンツとして、広く所蔵資料のデジタル画像が捉えられている状況であることを確認した。

利用者向けの機能では、辞書機能、サジェスト機能、ファセットによる絞り込み検索機能といった多様な検索支援手段が提供されていた。また、複数資料の重ね合わせやレスポンシブウェブデザインの採用など、ユーザーインターフェースの高機能化が見られた。

業務機能では、目録データ登録管理、画像データ管理、利用状況把握、閲覧業務関連に必要な機能

の提供状況等を把握した。

利用拡大のための機能・サービスでは、SNS 等ソーシャルメディアが一部機関で活用されていることを確認した。また、利用者からの提供情報（メタデータ・画像データ）を直接的・間接的に登録する事例としては、市民ボランティアによる目録データの整備事業や登録データの誤り等システム上の不具合を通報する機能の実装例を確認した。その他、デジタルコンテンツの展示・広報への活用事例、他機関との連携の事例を確認した。

利用サービス向上のための機能・サービスについては、レファレンスサービスの充実が利用者の増加につながる事例や、利用動向（システムログ）の解析サービスの事例を確認した。

(3)　アンケート及び調査結果の検討

アンケートへの回答で示されたサービスや調査したサービスについて、それぞれ項目別に分け、デジタルアーカイブ・システムの適用可能性について検討した。以下に項目別の検討状況を示す。

デジタルアーカイブの基本機能(利用者向け)	基本的な検索機能に加えて、辞書機能、サジェスト機能、ファセットによる絞り込み検索機能といった検索支援手段の整備により、利用者における利便性向上を図ることが重要であるが、実装においては、想定する利用者に合わせた検討を要する。また、マルチメディアコンテンツへの対応や、利用環境の充実を図る機能を整備することが必要である。
デジタルアーカイブの基本機能(業務)	汎用的なフォーマットで作成された目録データにより容易にシステムへの登録・更新が行えるユーザーインターフェースと機能、目録データを個別に登録・編集できる機能、画像データを容易にシステムへ登録・更新・削除が行えるユーザーインターフェースと機能が必要であるが、各機関におけるデータ項目や機能への要求レベルには差異がある。また、目録データやデジタルコンテンツに対する利用制限を行う機能、利用請求管理へのニーズが高い。各種業務に即した業務システムの整備が求められている。
利用者拡大のための機能・サービス	利用者がデータ登録や情報発信に寄与する機能が有効である。例えば、利用者が SNS を用いて情報発信を容易に行える仕組み、各機関が SNS を用いて資料の情報を利用者に発信する仕組みが効果的であるが、組織のセキュリティポリシーに準じた運用方針の明確化が必要である。また、デジタルアーカイブ・システムが提供する情報の充実を目的とした、利用者によるメタデータ付与等への能動的参加を実現するためには、利用者の作業環境実現における技術的な対応への考慮と利用者による作業の管理等運用における考慮を要すると考えられる。その他、資料情報の展示等への二次利用は、利用者拡大に有効と考えられる。
利用サービス向上のための機能・サービス	レファレンスサービスの充実のため、レファレンスサービスの実施を支援するデータベース構築が有用であること、単館運用よりも共同利用に適していることが分かったが、共同利用型システムでの運用にあたり、運用主体や費用分担が課題である。また、システムのログ解析による利用動向の把握は有効と考えられる。

2. デジタルアーカイブ・システムに関連する最新 IT 技術に係る調査検討

2.1. システム構成等に係る調査

2.1.1. システム構成要素

コンピュータシステムは、以下のような要素で構成されている。デジタルアーカイブ・システムに適用するシステム構成を検討するにあたり、(1)システム基盤、(2)アプリケーションのそれぞれについて、どのような類型があるかを整理する。

表 2 システム構成要素

<table>
<tr><td>アプリケーション
（目録検索機能、画像閲覧機能等）</td><td rowspan="2">アプリケーション</td></tr>
<tr><td>プラットフォーム
（サーバ、ストレージ、OS、ミドルウェア等）</td></tr>
<tr><td>ネットワーク
（回線・ネットワーク機器等）</td><td rowspan="3">インフラ
（システム基盤）</td></tr>
<tr><td>ファシリティ
（電源・空調・耐震設備等）</td></tr>
</table>

(1) システム基盤の類型

システム基盤は、システムリソースの利用形態、設置環境という2つの観点から、表 3の通り4つの類型に整理できる。

表 3 システム基盤の類型

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">リソースの利用形態</td></tr>
<tr><td>自組織で専有</td><td>必要な分を共有</td></tr>
<tr><td rowspan="3">設置環境</td><td>自組織の施設に設置</td><td>①オンプレミス型</td><td rowspan="2">③プライベートクラウド型</td></tr>
<tr><td rowspan="2">外部施設に設置</td><td rowspan="2">②コロケーション型</td></tr>
<tr><td>④パブリッククラウド型</td></tr>
</table>

各類型の特徴を以下にまとめる。

① オンプレミス型（自館設置）

サーバ機器、ストレージ機器、ネットワーク機器等のハードウェアを自組織で専有し、かつサーバ室等の設置場所を自組織内に用意する。従来からあるシステム構成である。

オンプレミス型での特徴、留意点は以下の通りである。

【特徴】

- システムの目的に応じたリソース設計を柔軟に行える。
- ネットワーク設計の自由度が高く、セキュリティ対策を行いやすい。
- 人件費を除く運用コストは、保守費用と設備費用(光熱費、回線費用等)のみであるため低く抑えられる場合がある。

【留意点】

- システム基盤とアプリケーションの要件を整理するために、高い専門知識が求められる場合がある。また、運用にあたって専門知識を持った担当者が必要となる。
- ハードウェアの保守期間や耐用年数を超えると更新が必要となる。
- ハードウェアの購入、外部回線の引き込み、専用設備のための施設改修など、初期コストが高くなる場合がある。

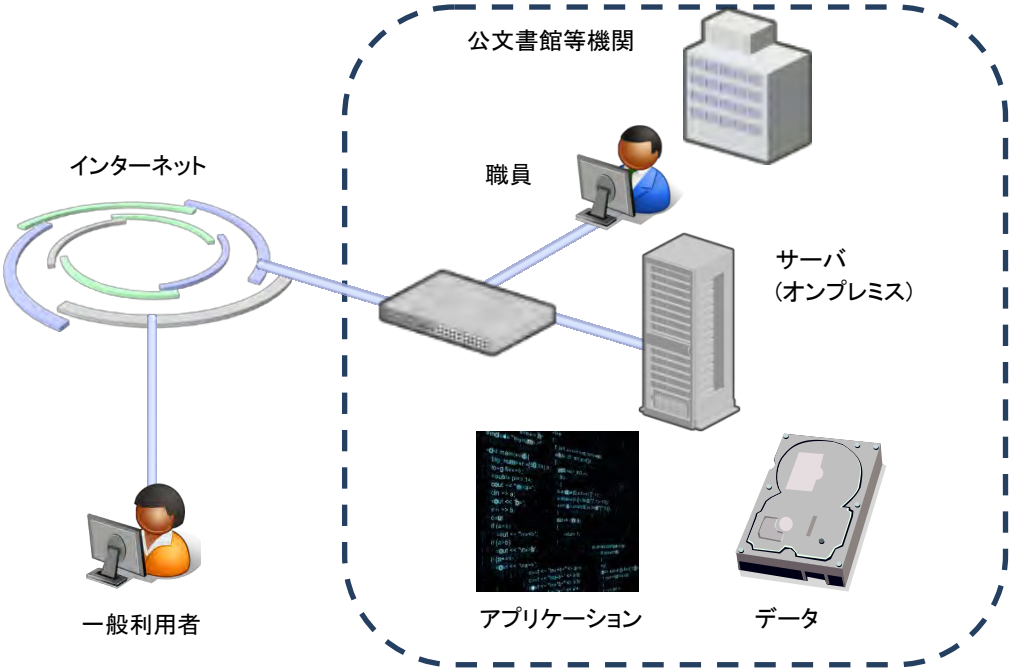


図 2 オンプレミス型のイメージ

② コロケーション型(データセンター設置)

サーバ機器、ストレージ機器、ネットワーク機器等のハードウェアを自組織で専有するが、外部施設であるデータセンターに設置し、データセンターサービス事業者からバックボーン回線やサーバーラック等の提供を受ける。2000 年頃からデータセンター事業者の増加とともに増えてきたシステム構成である。ハウジングサービスとも呼ばれる。

コロケーション型での特徴、留意点は以下の通りである。

【特徴】

- データセンターの安定した環境で、24 時間の運用監視体制が得られる。
- システムの目的に応じたリソース設計を柔軟に行える。
- ネットワーク設計の自由度が高く、セキュリティ対策を行いやすい。

【留意点】

- システム基盤とアプリケーションの要件を整理するために、高い専門知識が求められる場合がある。
- ハードウェアの保守期間や耐用年数を超えると更新が必要となる。
- ハードウェアの購入の初期コストがかかる。
- データセンター事業者によって提供されるサービス品質が異なり、またサービスレベルに応じて価格が高くなる。

③ プライベートクラウド型

プライベートクラウド型には、自組織の施設内に構築したクラウド環境と、外部事業者が提供するクラウドサービスの2つがある。外部事業者が提供するクラウドサービスの場合、専用の環境であるものをプライベートクラウドと呼ぶ。

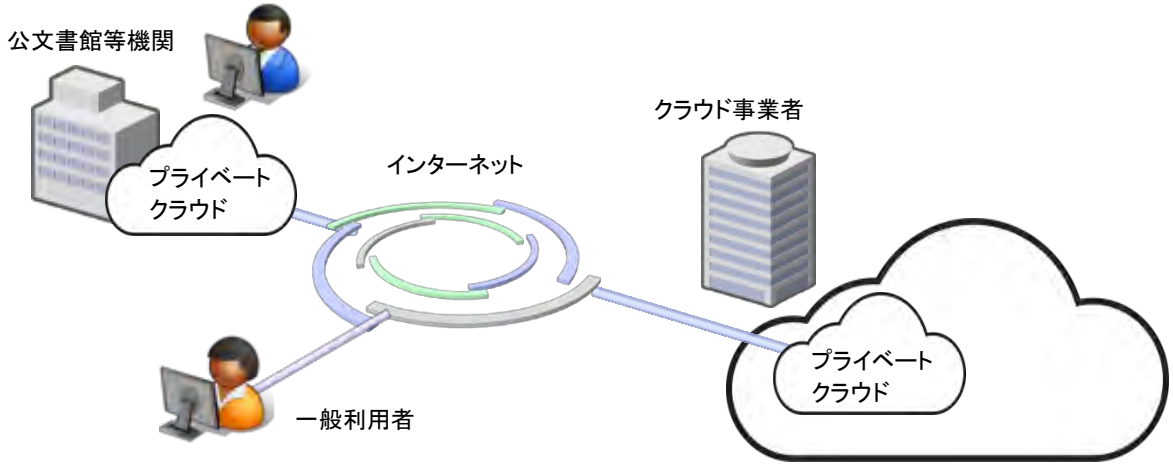


図 3 プライベートクラウド型のイメージ

自組織内に構築したクラウド環境は、物理的にも専用であるのに対して、外部事業者が提供するクラウドサービスは、一般には論理的に専用の環境である。

プライベートクラウド型での特徴、留意点は以下の通りである。

【特徴】

- 余剰リソースを活用できるため、全体としての効率化が図れる。
- 自組織内の基準に則ったセキュリティを実現できる。
- インフラ運用担当者が不要となるため、人件費を削減できる。
- (外部事業者が提供する場合)必要なリソースのみの利用料が発生し、初期投資が抑えられる。

【留意点】

- リソース変更の柔軟性が低い。
- パブリッククラウドと比較するとコストが高い。
- (外部事業者が提供する場合)提供されるサービス品質が異なり、またサービスレベルに応じて価格が高くなる。

④ パブリッククラウド型

外部事業者が提供するクラウドサービスのリソースを、必要に応じて利用する。大規模なデータセンターを有する事業者が増加し、仮想化・分散化技術と組み合わせることで増えてきたシステム構成である。

パブリッククラウド型での特徴、留意点は以下の通りである。

【特徴】

- インフラ調達にかかる手間、時間が少なく済む。
- 必要なリソースの利用料が発生し、初期投資が抑えられる。
- リソースの拡張が容易に行える。
- システム基盤の管理は外部事業者が行うため、運用に代わる人件費を削減できる。

【留意点】

- 不特定の組織が共有する環境のため、セキュリティ対策に不十分な点がある場合には、データ保護が徹底できないリスクがある。
- 障害時の対応等について、各機関ではコントロールできない。
- 外部事業者によって提供されるサービス品質が異なり、またサービスレベルに応じて価格が高くなる。

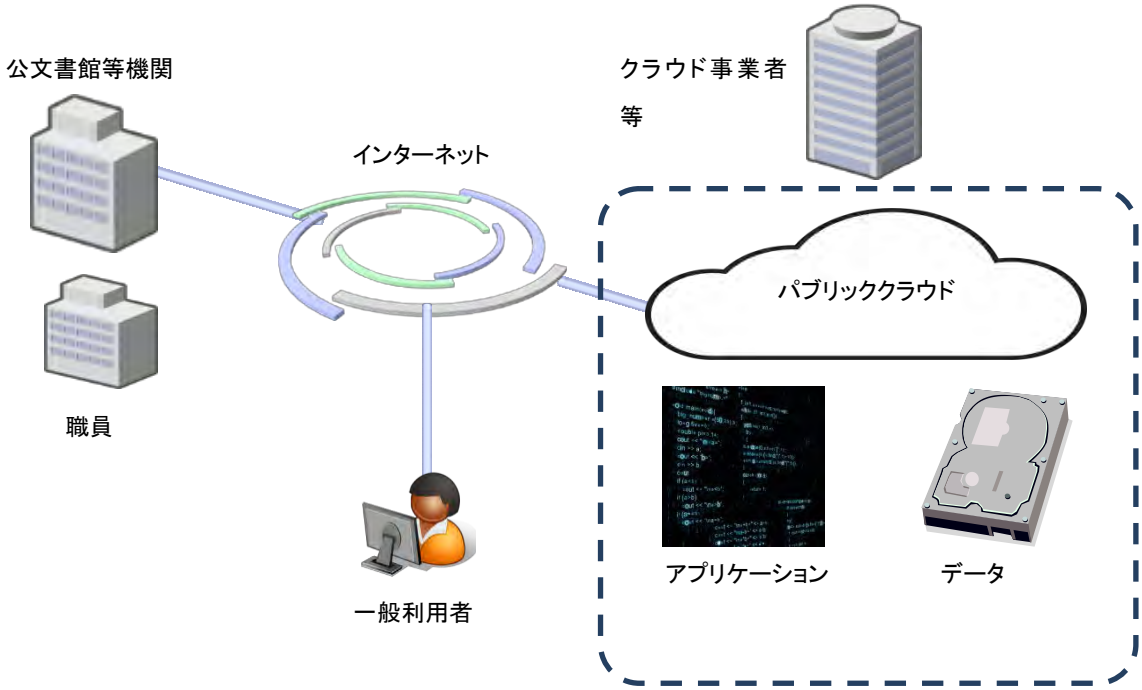


図 4 パブリッククラウド型のイメージ

(2)　アプリケーション提供方法の類型

システム基盤上で稼働するアプリケーションは、提供方法により 3 つの類型に整理できる。

① 開発・導入型

必要なアプリケーションを新規に開発する、またはパッケージソフトウェアにカスタマイズを行う。

開発・導入型での特徴、留意点は以下の通りである。

【特徴】

- 必要な機能を予算の範囲で柔軟に実現できる。
- 各機関の業務、運用に合わせたアプリケーションを構築できる。

【留意点】

- 開発・カスタマイズする機能の要件について、ソフトウェアベンダーとの綿密な擦りあわせが必要となる。
- コストは大きく、開発期間は長くなる傾向がある。
- 開発したソフトウェアベンダー以外は運用保守を行うことが困難である。

②パッケージ導入型

ソフトウェアベンダー等が提供するパッケージソフトウェアを、最低限の設定のみで導入する。

パッケージ導入型での特徴、留意点は以下の通りである。

【特徴】

- パッケージソフトウェアの標準化された機能により、業務の効率化が期待できる。
- 開発コストを軽減できる可能性が高く、リリースまでの期間を短縮できる。

【留意点】

- パッケージソフトウェアが想定していない業務がある場合は、業務を標準的な内容に変更する必要があるため、一時的にコストが発生する。

③サービス利用型

ソフトウェアベンダー等が提供するサービスを利用する。ASP(Application Service Provider)、SaaS(Software as a Service)と呼ばれる形態である。

サービス利用型での特徴、留意点は以下の通りである。

【特徴】

- 標準化されたサービスにより、業務の効率化が期待できる。
- セキュリティ対策やアプリケーションの維持・運用にかかる作業はサービス提供者が行うため、維持・運用にかかる人員を他の業務に振り分けることができる。
- 月額のサービス利用料のみとなるため初期コストが軽減できる。
- 利用開始までの期間が短い。

【留意点】

- 提供されるサービスが想定していない業務がある場合は、業務を標準的な内容に変更するため、一時的にコストが発生する。

(3)　まとめ

システム基盤、アプリケーション提供方法という 2 つの観点で類型を示した。

システム基盤では、オンプレミス型が構成の自由度は高いが、初期コストと運用負担が高くなる。パブリッククラウド型では制約は増えるが、初期コスト及び運用コスト、運用負担を抑えることができる。プライベートクラウドは、パブリッククラウドよりもコスト及び構成の自由度共に高くなる。コロケーション型はそれらの中間に位置する。

アプリケーション提供方法では、開発・導入型が要求機能の実現性は最も高いが、初期コストが高い。サービス利用型は、提供されるサービスを利用するのみであるため、初期コストが抑えられ、導入までの期間も短くなる。パッケージ導入型はその中間に位置する。パッケージ導入型やサービス利用型においては、業務を提供されるソフトウェアやサービスに揃えるために、一時的に時間や費用が発生するが、長期的には業務の効率化や利用サービスの向上に寄与する可能性がある。システム導入にあたっては、長期的な視野でコスト以外の側面も検討する必要がある。

構成の自由度・要求機能の実現性を縦軸に、コストを横軸に取った場合の各構成の配置は図 5 の通りとなる。

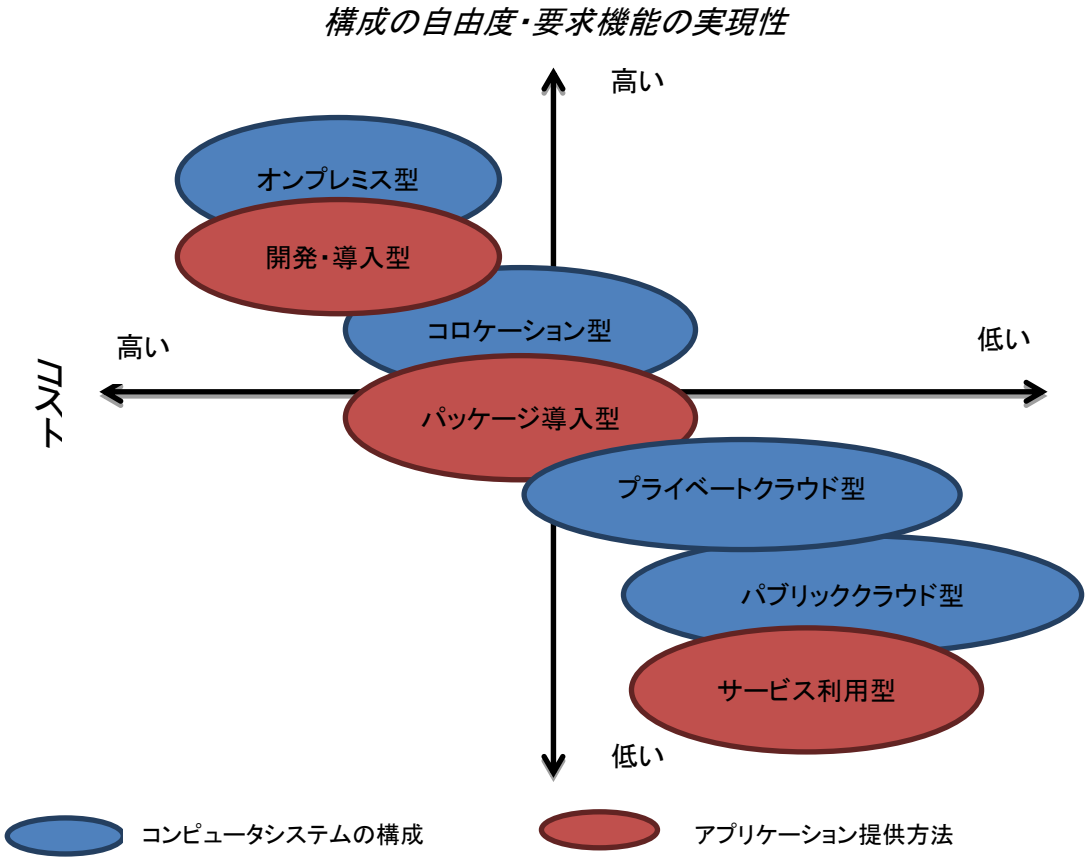


図 5 構成の自由度・要求機能の実現性とコストによる各類型の配置

2.1.2. 運用形態

本項では、システムの管理主体を中心とした運用形態について整理する。運用形態は、システム管理主体により大きく単館運用と共同運用に分けられる。

(1) 単館運用

一つの機関が単独でシステムを構築・運用する形態である。単館運用での特徴、留意点は以下の通りである。

【特徴】

- いずれのシステム基盤、アプリケーション提供方法を採用するか等、各機関の事情に合わせた検討や変更が可能である。
- 各機関の業務運用に合わせた運用計画を立てやすい。

【留意点】

- 構築・運用コストは各機関の予算で賄う必要がある。

(2) 共同運用

共同運用は、複数の機関が共同してシステムを構築・運用する形態である。他機関が用意したインフラやアプリケーションを利用するのみの場合や、アプリケーションの管理を分担しておこなう場合など、共同運用の形態は多数ある。共同運用の形態によっては、利用者の利便性向上と、費用や人員等の負担軽減を両立することができる。

共同運用の形態は、リソースを共有する類型のシステム基盤毎に分けて考えることができる。以下にそれぞれにおいて想定される運用形態を(ア)から(エ)に分けて述べる。

		アプリケーションの提供方法	
		開発・導入型 パッケージ導入型	サービス利用型
システム基盤	プライベートクラウド	(ア)	(イ)
	パブリッククラウド	(ウ)	(エ)

（ア）プライベートクラウドによるシステム基盤・開発・導入型／パッケージ導入型

システム基盤： プライベートクラウド

アプリケーションの提供方法： 開発・導入型／パッケージ導入型

システム基盤の管理主体	プライベートクラウドを構築又は契約した組織が、システム基盤の管理主体となる。 外部事業者が提供するプライベートクラウドの場合には、プライベートクラウドの運用はクラウド事業者がおこなう。
アプリケーションの管理主体	アプリケーションを導入した組織が、アプリケーションの管理主体となる。

この運用形態では、共同運用する機関がそれぞれアプリケーションを構築する場合と、共通のアプリケーションを利用する場合が考えられる。具体的には、以下の例が挙げられる。

例 1）

A 館、B 館がそれぞれアプリケーションを構築する。プライベートクラウドの契約は A 館 B 館どちらかの場合もあるが、別の組織の場合もある。

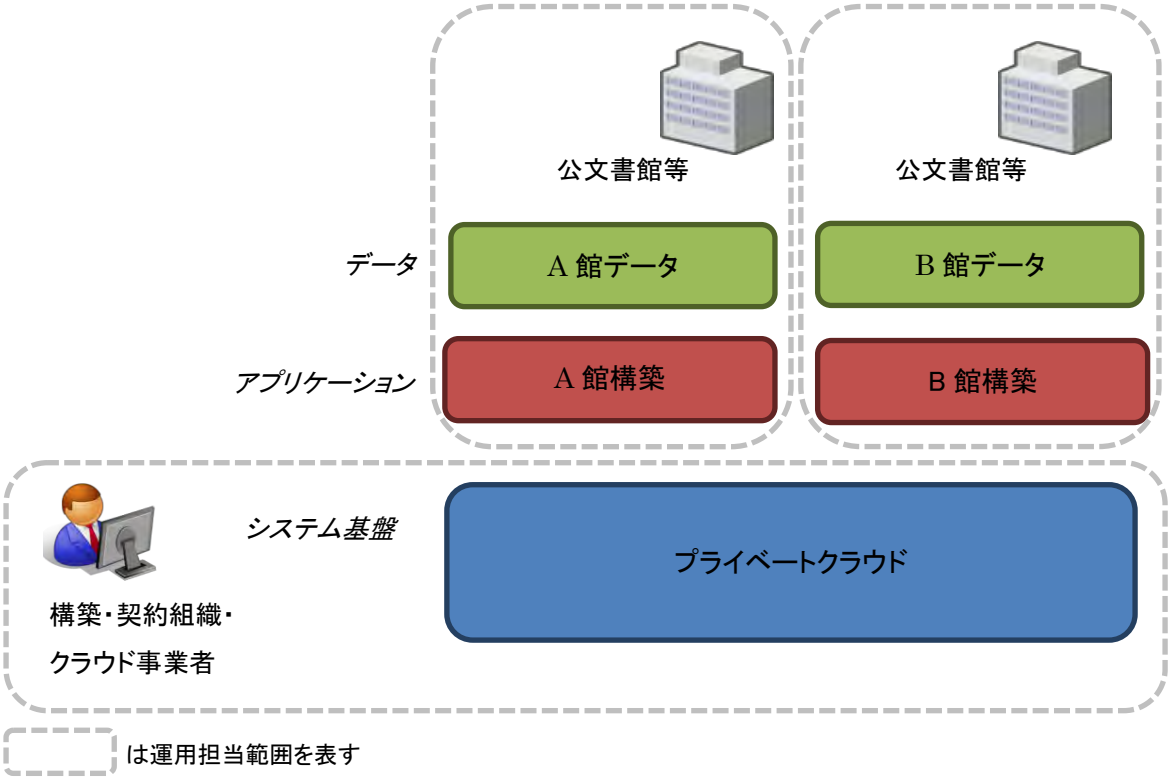


図 6 プライベートクラウドによるシステム基盤・開発・導入型／パッケージ導入型 例 1 の構成イメージ

例 2）

A 館がアプリケーションを構築し、A 館と B 館が利用する。プライベートクラウドの契約は A 館 B 館どちらかの場合もあれば、別の組織の場合もある。

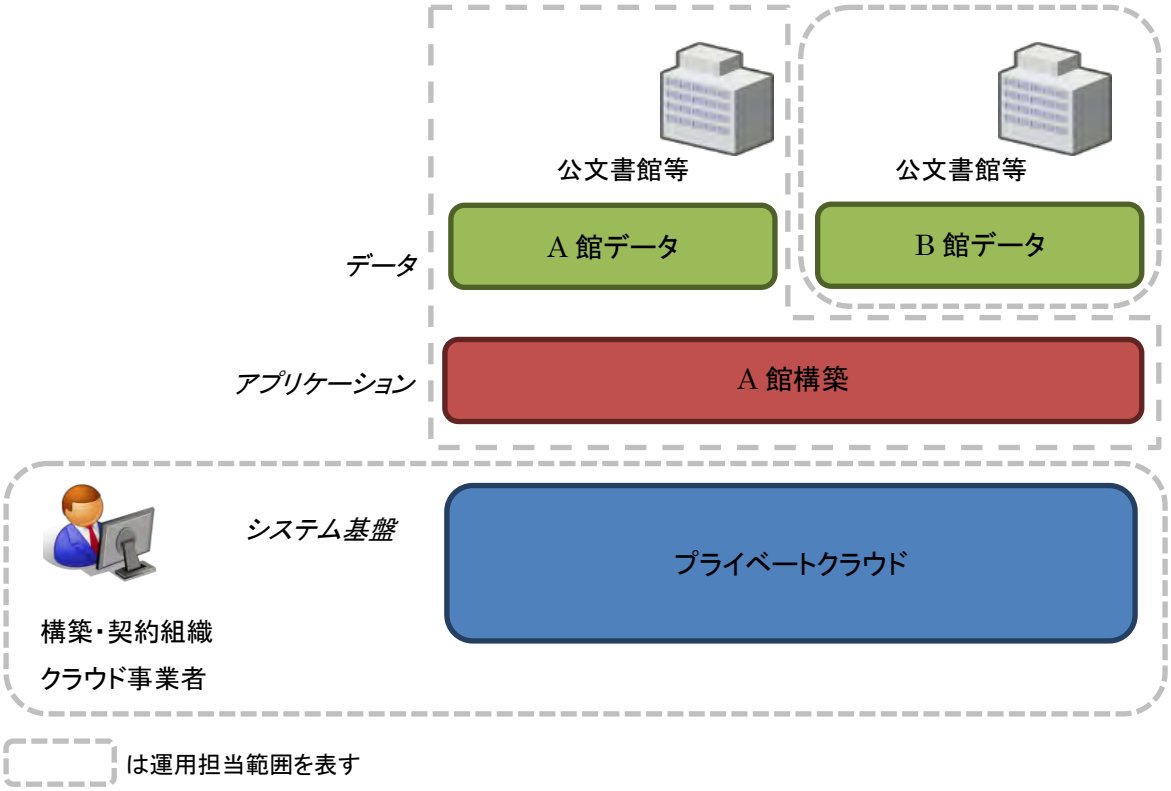


図 7 プライベートクラウドによるシステム基盤・開発・導入型／パッケージ導入型 例 2 の構成イメージ

（イ）プライベートクラウドによるシステム基盤・サービス利用型

システム基盤： プライベートクラウド

アプリケーションの提供方法： サービス利用型

システム基盤の管理主体	プライベートクラウドを構築又は契約した組織が、システム基盤の管理主体となる。 外部事業者が提供するプライベートクラウドの場合には、プライベートクラウドの運用はクラウド事業者がおこなう。
アプリケーションの管理主体	サービス提供者がアプリケーションの管理主体となる。

この運用形態では、共同運用する機関が共通のアプリケーションを利用する。具体的には、以下の例が挙げられる。

例）

A 館、B 館は共にサービスを利用するのみである。

アプリケーションは、サービス提供者（プライベートクラウドの事業者や、第三者機関等）が管理する。

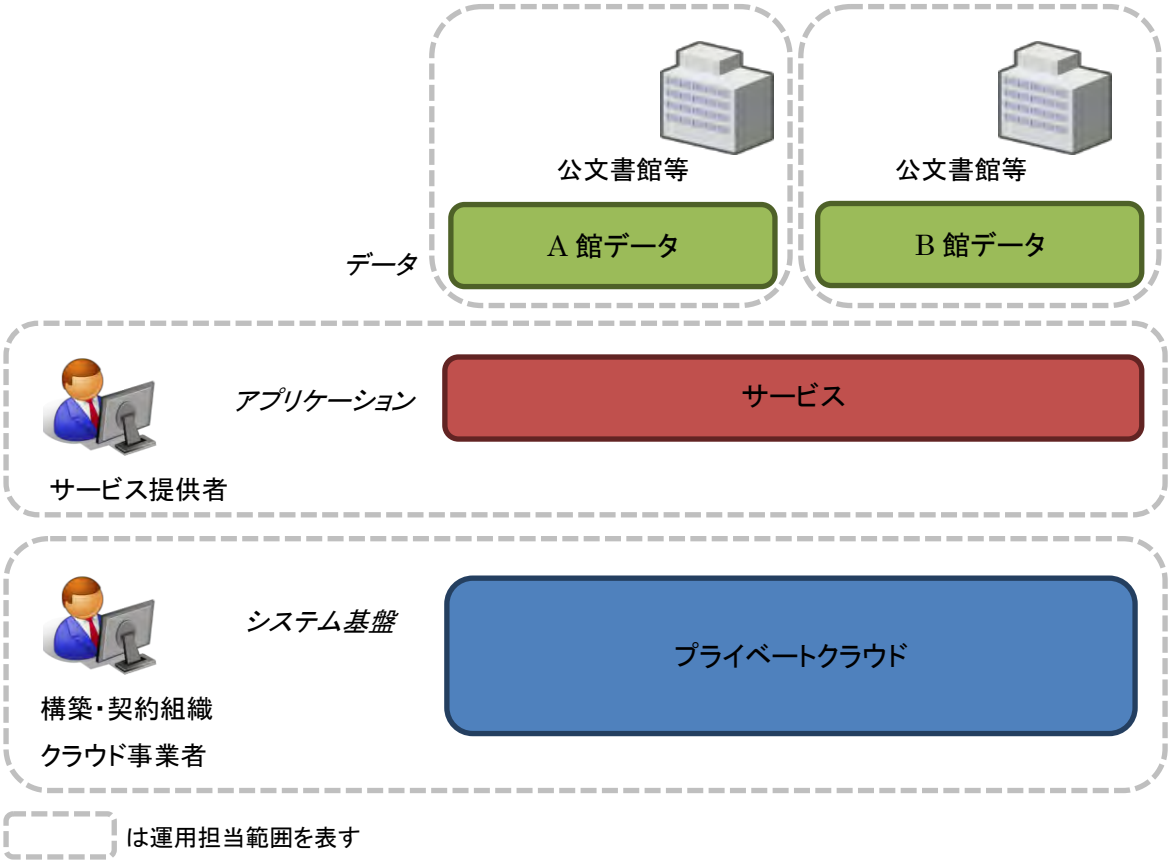


図 8 プライベートクラウドによるシステム基盤・サービス利用型の構成イメージ

（ウ）パブリッククラウドによるシステム基盤・開発・導入型／パッケージ導入型

システム基盤： パブリッククラウド

アプリケーションの提供方法： 開発・導入型／パッケージ導入型

システム基盤の管理主体	パブリッククラウドのサービス提供事業者が、システム基盤の管理主体となる。
アプリケーションの管理主体	アプリケーションを導入した組織が、アプリケーションの管理主体となる。

この運用形態では、共同運用する機関がそれぞれアプリケーションを構築する場合と、共通のアプリケーションを利用する場合が考えられる。具体的には、以下の例が挙げられる。

例）

A 館がアプリケーションを構築し、その一部を A 館と B 館が利用する。この場合、A 館がアプリケーションを管理する。

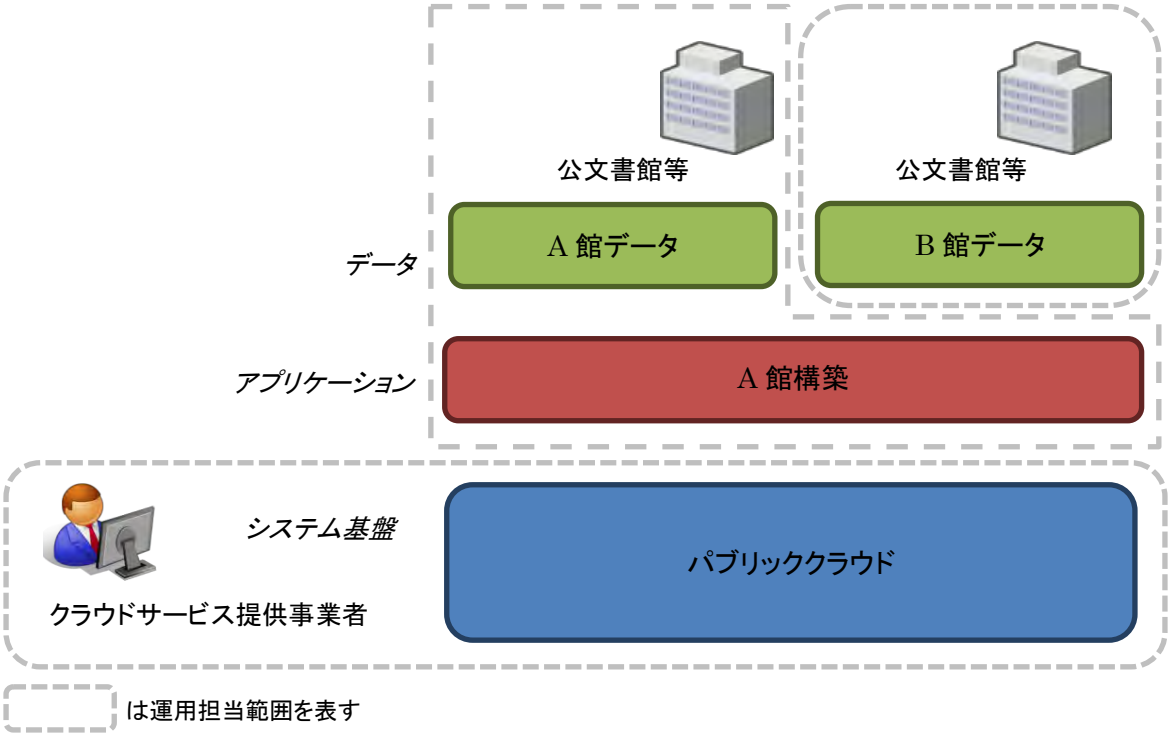


図 9 パブリッククラウドによるシステム基盤・開発・導入型／パッケージ導入型の構成イメージ

（エ）パブリッククラウドによるシステム基盤・サービス利用型

システム基盤： パブリッククラウド

アプリケーションの提供方法： サービス利用型

システム基盤の管理主体	パブリッククラウドのサービス提供事業者が、システム基盤の管理主体となる。
アプリケーションの管理主体	サービス提供者がアプリケーションの管理主体となる。

この運用形態では、共同運用する機関が共通のアプリケーションを利用する。具体的には、以下の例が挙げられる。

例）

A 館、B 館は共にサービスを利用するのみである。

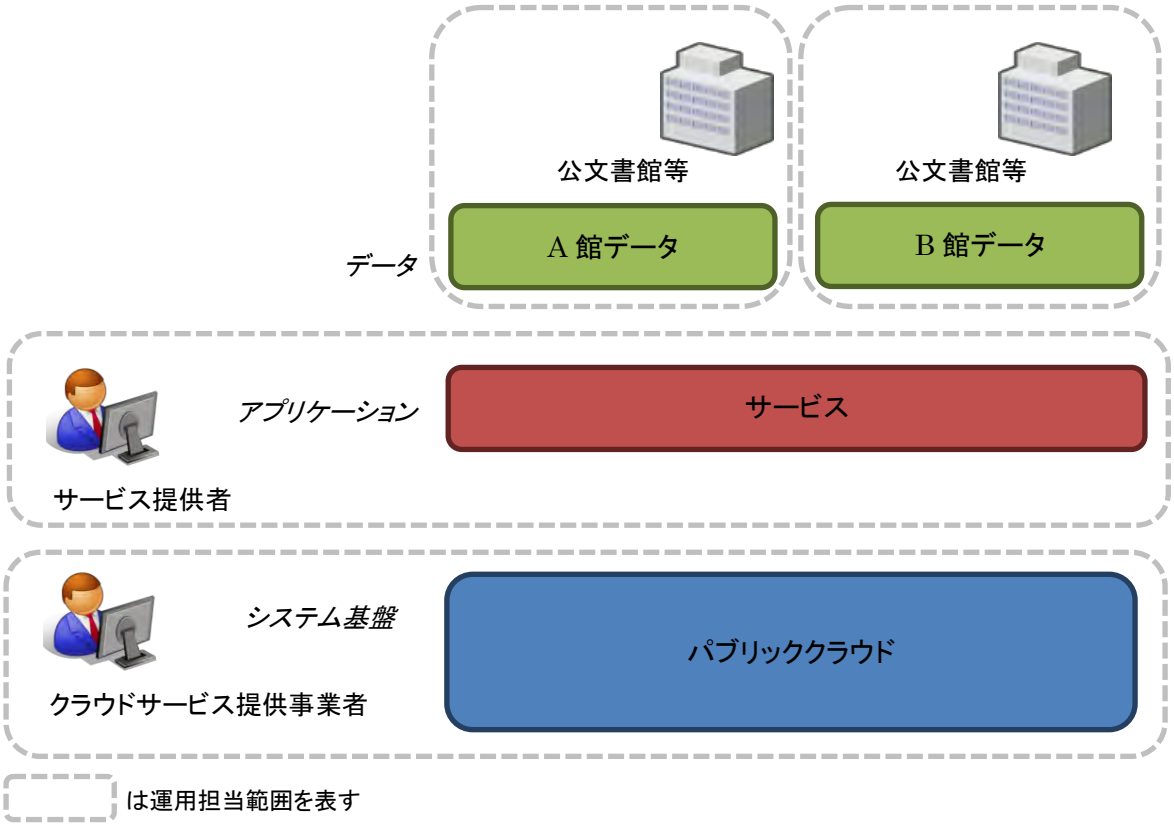


図 10 プライベートクラウドによるシステム基盤・サービス利用型の構成イメージ

(3) まとめ

単館運用では、いずれのシステム基盤、アプリケーション提供方法を採用するか等、各機関の事情に合わせた検討や変更が可能である。アプリケーションを各機関の業務や利用者の要求にあわせたカスタマイズなど柔軟な対応が可能であるが、費用や人員を各機関で確保する必要がある。

共同運用には複数の類型があるため、以下に特徴を整理する。

表 4 共同運用の類型と特徴

	開発・導入型 パッケージ導入型	サービス利用型
プライベートクラウド	【特徴】 リソースを有効活用しつつも、専用の環境であるため、ポリシーに沿ったセキュリティを実現できる等、ある程度柔軟にシステム基盤の設計が可能である。 また、各機関が個別にアプリケーションを構築することや、業務管理機能を備えることが可能である。共同運用の相手機関が導入するアプリケーションを利用するのみの場合でも、機能要件を事前に調整することが可能である。 【留意点】 パブリッククラウドと比較するとコストは高い。 システム基盤の設計やリソースの配分、構築するアプリケーション等に関し、共同運用の相手機関と調整が必要な場合がある。	【特徴】 データのみ用意すれば、比較的低コストで提供者されるサービスを利用することができる。 専用のクラウド環境であるため、サービスの利用者が限定されており、パブリッククラウドよりも、セキュリティ要件に合う対策が行われている可能性が高い。 【留意点】 システムの稼働時間やデータ容量の制限等、提供されるサービスに合わせる必要がある。 サービス提供者の選択肢がほとんどない。
パブリッククラウド	【特徴】 システム基盤の運用管理が不要で、コストを抑えることができる。 また、各機関が個別にアプリケーションを構築することが可能である。共同運用の相手機関が導入するアプリケーションを利用するのみの場合でも、機能要件を事前に調整することが可能である。 【留意点】 構築するアプリケーション等に関し、共同運用する組織と調整が必要な場合がある。 サービス事業者のセキュリティポリシーに従うため、公開データのみを取り扱うことになると考えられる。	【特徴】 データのみ用意すれば、低コストで提供されるサービスを利用することができる。 各機関では構築が困難な高度な機能等が利用でき、長期的には業務負担の軽減が期待できる。 【留意点】 システムの稼働時間やデータ容量の制限等、提供されるサービスに合わせる必要がある。 提供されるサービスによるが、基本的には公開データのみを取り扱うことになると考えられる。

2.1.3. デジタルアーカイブ・システムへの適用

システム構成要素と運用形態に示した類型を元に、デジタルアーカイブ・システムへの適用について検討する。

(1) デジタルアーカイブ・システムの特徴

デジタルアーカイブ・システムは以下の特徴を持ったシステムである。

- 公文書館等が所蔵する公文書等の目録情報を提供する。
- インターネットを通じて、不特定多数の利用者に情報提供する。
- ２４時間いつでも利用できる。
- 公開する目録情報等は、長期安定利用のために永続してアクセスできる。
- 公開する目録情報等は、公文書を基本とするが、古文書等の資料についても対応する。
- 目録情報と関連づけて、資料をデジタル化したデジタルコンテンツデータを提供する。
- 一般利用者には非公開とする目録情報や目録項目、デジタルコンテンツデータがありうる。
- アクセス数は各機関の規模により異なるが、1 日あたり数百件から数万件程度と想定される。
- ネットワーク上に流れるデータ量は各機関の公開するデータ件数により異なるが、デジタルコンテンツの件数に比例して増加する。

(2)　デジタルアーカイブ・システムの機能要件

デジタルアーカイブ・システムは機能要件の側面から見ると、以下のように分類することができる。

表 5 機能要件の分類と考え方

機能要件の分類	機能要件の考え方
基本的な機能要件	デジタルアーカイブ・システムとしての必要最低限の機能を実現するケース。 必要最低限の機能としては、目録情報の提供が想定される。
標準的な機能要件	デジタルアーカイブ・システムの推奨機能を除いた、標準機能を実現するケース。 標準機能としては、例えば目録情報に加えて、資料のデジタル画像の提供が想定される。
発展的な機能要件	デジタルアーカイブ・システムの推奨機能を含めた機能を実現するケース。 推奨機能としては、例えば動画や音声の配信、検索支援機能、業務支援機能等が想定される。

デジタルアーカイブ・システムにおける推奨機能を含む機能は、一例では以下のような機能が考えられる。

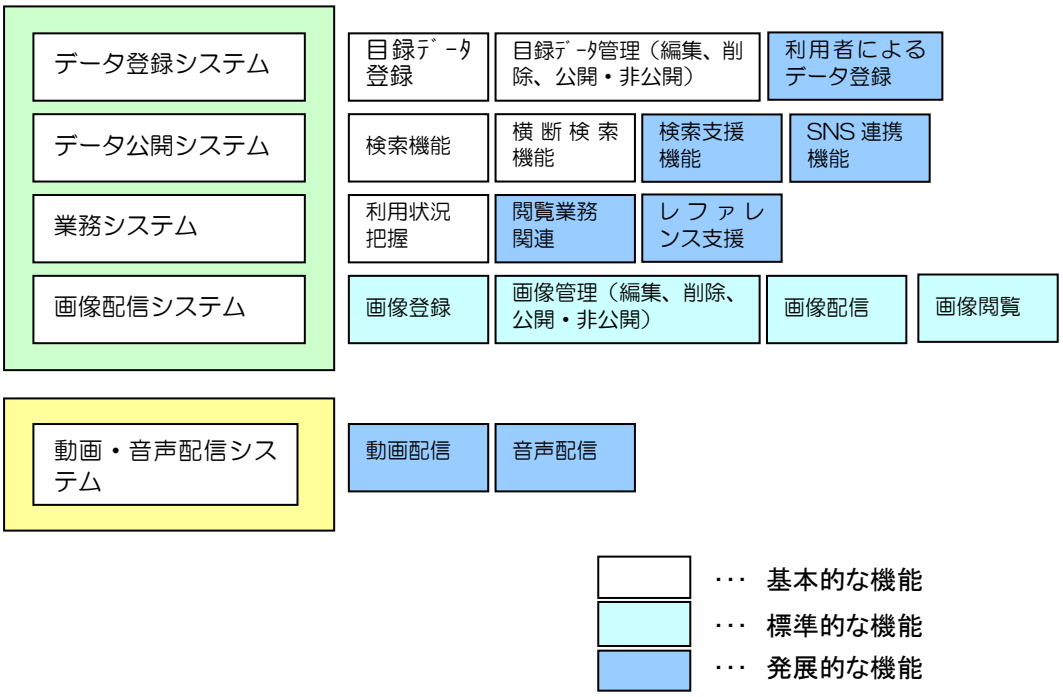


図 11 デジタルアーカイブ・システムの推奨機能を含む機能例

この機能要件の分類に基づき、インフラ要件と機能要件を組み合わせたシステム構成が考えられる。インフラ要件と機能要件の組み合わせによるパターンは、以下の表で表現できる。

表 6 インフラ要件と機能要件を組み合わせたシステム構成のパターン

		インフラ基盤を自組織で構築			外部事業者が提供するインフラ基盤を利用	
		オンプレミス	コロケーション	プライベートクラウド	プライベートクラウド	パブリッククラウド
		①	②	③	④	⑤
機能要件	基本的（Ⅰ）	①-Ⅰ	②-Ⅰ	③-Ⅰ	④-Ⅰ	⑤-Ⅰ
	標準的（Ⅱ）	①-Ⅱ	②-Ⅱ	③-Ⅱ	④-Ⅱ	⑤-Ⅱ
	発展的（Ⅲ）	①-Ⅲ	②-Ⅲ	③-Ⅲ	④-Ⅲ	⑤-Ⅲ

単館運用　　単館・共同運用

例えば、オンプレミス型で標準的な機能要件の場合（①-Ⅱ）は、システム構成は以下のようなイメージとなる。

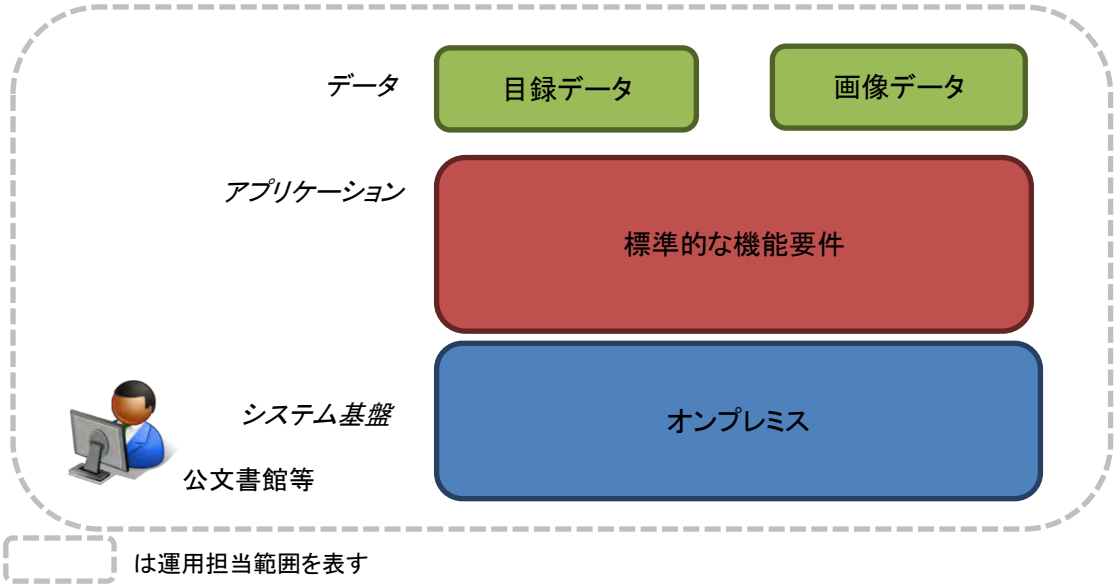


図 12 オンプレミス型で標準的な機能要件の場合（①-Ⅱ）の構成イメージ

また運用形態を踏まえると、共同運用の場合には、各機関が同じアプリケーションを使用する場合（図7参照）と、異なるアプリケーションを使用する場合（図6参照）が可能である。

以下は外部事業者が提供するプライベートクラウドに、A 館が基本的な機能要件でアプリケーションを構築し、B 館が標準的な機能要件でアプリケーションを構築した場合の例である。

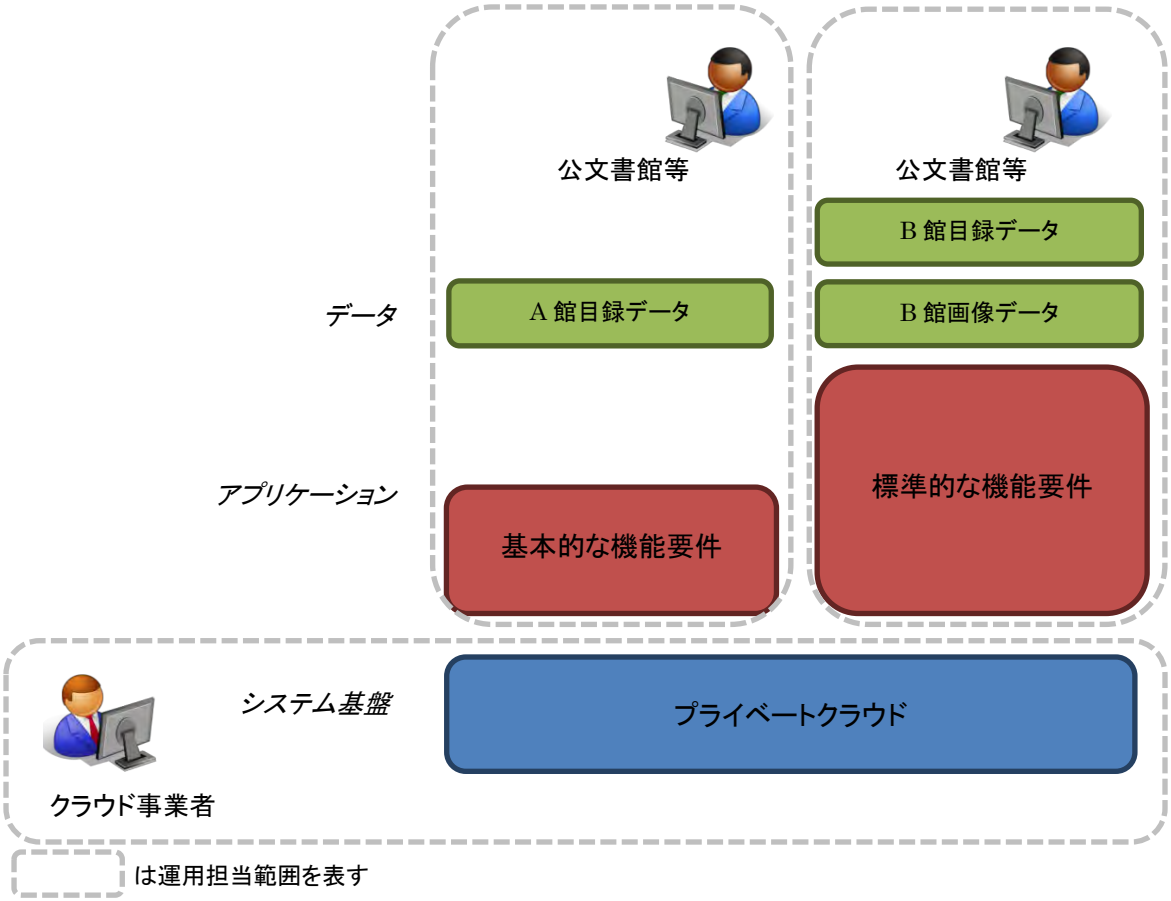


図 13 プライベートクラウドⅡ型（④-Ⅰ、④-Ⅱ）の構成イメージ

表6に記載した各パターンの特徴を以下に記載する。

オンプレミスとコロケーションは、システムの設置場所に違いはあるものの、インフラを自組織で構築し単館で運用するという点では大きな違いはない。よってシステム構成としては、基本的要件の「①-Ⅰと②-Ⅰ」、標準的要件の「①-Ⅱと②-Ⅱ」、発展的要件の「①-Ⅲと②-Ⅲ」の3つにまとめることができる。これらのシステム構成では、費用や人員を各機関で確保する必要があるものの、業務や利用者の要求にあわせたアプリケーションを構築可能である。

次に、インフラ基盤を自組織で構築するプライベートクラウド（③-Ⅰ、③-Ⅱ、③-Ⅲ）は、各機関が単独でクラウド環境を構築することも可能性としては考えられるが、同一組織の複数機関が利用するクラウド環境として構築する場合に、運用面や費用面での効果が現れる。クラウド技術を用いてはいるが、基本的には、オンプレミスやコロケーションの延長線で考えることができ、自組織のセキュリティポリシー等にあった環境において、組織全体では効率的にリソースを活用することが可能となる。また、インフラ運用の担当部門を組織でひとつに集約することができるため、各機関が専任のインフラ運用担当者を置く必要がなくなる。一例では、自治体の情報システム部門が構築したインフラ基盤を、自治体内の複数機関

が利用する場合などが考えられる。また、例えば発展的な構成（①-Ⅲと②-Ⅲ）において、別の部門等が管理する文書管理システムとデータ連携したり、公文書館以外の類縁機関と共通のアプリケーションを利用者に提供したり等、共同運用であることを活かした付加価値を提供することが可能となる。ただし、共同運用の形態によっては、参加する複数機関において、各種の調整等が発生すると考えられ、これらをコントロールする体制を明確にする必要がある。

外部事業者が提供するプライベートクラウド（④-Ⅰ、④-Ⅱ、④-Ⅲ）では、提供されるサービスや品質は事業者によって異なる。自組織で構築するクラウド（③-Ⅰ、③-Ⅱ、③-Ⅲ）と比べて、提供されるサービスに制約等がある可能性はあるものの、組織内にインフラ運用のための担当が不要となる。一方で、自組織専用のクラウド環境ではあるが、あくまで仮想的に区切られた外部環境であることに留意が必要である。例えば、非公開のデータ等をデジタルアーカイブ・システムで扱うかどうかは、提供されるサービスのセキュリティ対策等、サービス内容を検討する必要がある。

パブリッククラウド（⑤-Ⅰ、⑤-Ⅱ、⑤-Ⅲ）の場合、インフラの費用を抑えられるという特徴はあるが、留意すべき点は、外部事業者が提供するプライベートクラウド（④-Ⅰ、④-Ⅱ、④-Ⅲ）と概ね同様である。更に仮想的にも専用の環境ではないため、データの保護には一層の配慮が必要となる。一般的には、取り扱うデータは、公開データに限ることが基本となると考えられる。従って、基本的な構成（⑤-Ⅰ）、標準的な構成（⑤-Ⅱ）においても、各機関が管理している目録データや画像データのうち、一般公開が可能な部分のみを抽出してシステムに搭載することが考えられる。また発展的な構成（⑤-Ⅲ）でも、業務機能をはじめとして、実現できる機能が大きく制約を受けることが考えられる。デジタルアーカイブ・システムを一般公開のためのシステムと位置づけ、非公開データの管理等は別の手段で対応することにより、パブリッククラウドは有効な選択肢となりうる。

なお外部事業者が提供するインフラ基盤である④⑤においては、アプリケーションの提供方法としてサービス利用型がある。機能要件という観点では、開発・導入型やパッケージ導入型と同様の観点で考えればよいが、各機関の要求による機能拡張ができない点に留意が必要となる。多数の機関が利用することができるよう、標準化された機能が提供されているが、各機関の業務が特殊である場合、提供される機能・サービスに合わせて、業務フローやデータ等を標準化するように変更しなくてはならない場合も考えられる。変更には一定の時間や費用がかかると見込まれるが、各機関における業務の標準化が推進されることとなり、業務の効率化等にもつながる可能性が期待される。

(3)　まとめ

単館運用の場合、各機関の業務や利用者の要求にあわせたアプリケーションを構築可能であるが、費用や人員を各機関で確保する必要がある。その点ではオンプレミス型（自館設置）、コロケーション型（データセンター設置）で大きな違いはない。ただしデジタルアーカイブ・システムが、インターネットを通じて 24 時間サービス提供することを前提とすると、自機関で十分な設備や体制が取れない場合には、安定した電源や回線等を備え、運用担当者が 24 時間体制で監視をするデータセンターを利用することがより望ましい。

共同運用の場合、デジタルアーカイブ・システムを共同運用する理由は、主に利用者の利便性向上と、費用や人員等の負担軽減である。

利用者の利便性向上という点からは、単にシステム基盤の共有だけではなく、統一的なユーザーインターフェースやデータ連携等が求められる。このような、共同運用であることを活かした付加価値のある機能を提供する目的からは、専用のクラウド環境であるプライベートクラウドにおける、開発・導入型のア

プリケーションの方が、発展的な機能要件にも柔軟に対応することが可能である。

一方、費用や人員等の負担軽減という点からは、サービス利用型が有効な選択肢である。データを準備するのみで、利用者へのインターネットを通した情報提供が可能となる。サービス利用型ではパブリッククラウドの方が、提供されているサービスの選択肢も多く、コストメリットが現れやすい。

2.2. クラウド等 IT 技術に係る調査

2.2.1. クラウド技術

クラウド技術は、仮想化や分散化等の技術を組み合わせ、ネットワークを通じてコンピュータ資源を利用するものである。また、クラウド化技術は、共同運用型のデジタルアーカイブ・システムを構築する際に有用な技術である。本項では、クラウド環境を構築するクラウドインフラストラクチャソフトウェアと、SaaS、PaaS、IaaS 等のクラウド上のサービス提供形態の動向や事例について調査する。

(1) クラウド化技術

クラウド環境を構築し管理するソフトウェアは、クラウドインフラストラクチャソフトウェアと呼ばれる。クラウドインフラストラクチャソフトウェアは、物理的なサーバやストレージ、ネットワークインタフェース等ハードウェアを一元集約し、均質で論理的なサーバ等によるシステム基盤(クラウド基盤、クラウドインフラストラクチャ)を提供するソフトウェアである。また、このようなソフトウェアには、OpenStack[1]やCloudStack[2]といった、オープンソースとして提供されているものもある。

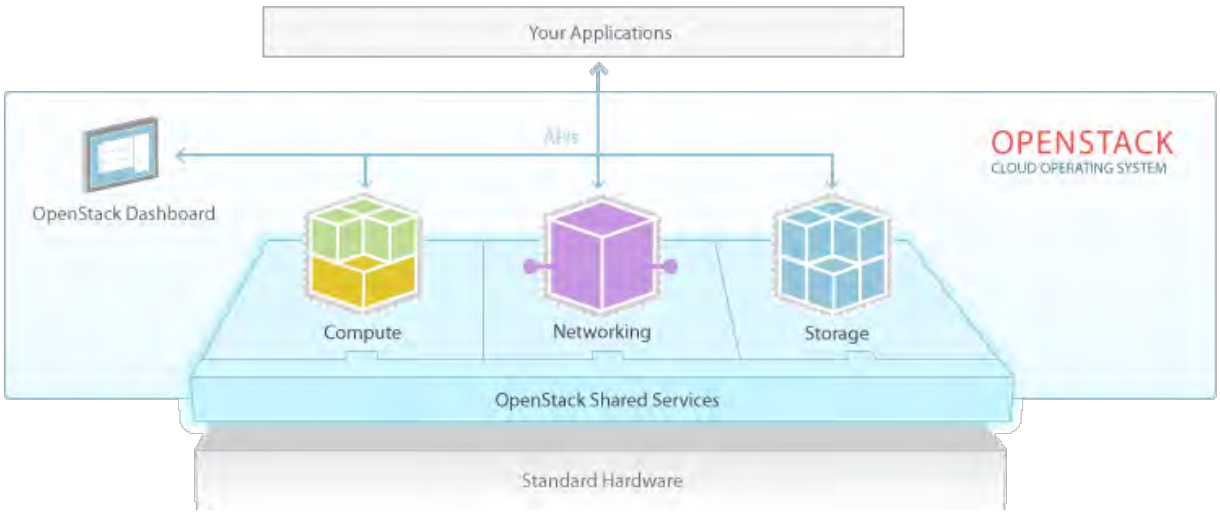


図 14 OpenStack の概要図[3]

仮想化技術により、一つのシステム基盤上に複数のシステムを実現することが可能である。また、このような状況において、ハードウェアの保守等作業は、システム基盤上に実現した個々のシステムに対してではなく、システム基盤を構成するハードウェアに対して発生することになる。いずれのソフトウェアも仮想マシンの構成管理、API の提供、大規模システムに対応したスケーラビリティ等を有している。

(2) クラウドサービス提供形態

現時点での主なクラウドサービス提供形態には以下の 3 つがある。

---

[1] http://www.openstack.org/ (2017 年 1 月 24 日アクセス)
[2] https://cloudstack.apache.org/ (2017 年 1 月 24 日アクセス)
[3] http://www.openstack.org/software/ (2017 年 1 月 24 日アクセス)

表 7 クラウドサービス提供形態

IaaS (Infrastructure as a Service)	CPU やメモリなどのハードウェア、ネットワークなどのインフラをインターネット経由で提供するサービス。
PaaS (Platform as a Service)	インフラに加えて OS やミドルウェアまでをインターネット経由で提供するサービス。
SaaS:(Software as a Service)	アプリケーションソフトウェアの機能をインターネット経由で提供するサービス。

調査機関による調査結果では、クラウド基盤サービス市場規模は 2015 年に前年比 39.1%増と大幅な成長を遂げており[4]、今後もクラウド化が進むことが予想される。

IaaS/PaaS 型のクラウドサービスは、従量課金制としている事業者が多く、追加費用はかかるが必要に応じてリソースの追加を容易に行える。一方、SaaS 型のクラウドサービスは、アプリケーションの「機能」を提供するという形態のため、月額固定が多い。

SaaS 型のクラウドサービスでは、デジタルアーカイブ・システムへ適用可能なサービスが複数存在している。提供事業者ごとに特徴を持ったサービスとなっているため、各機関の要求に最も適したものを選択することが可能であり、既存のサービスを活用する形態であることから、比較的短期間でデジタルアーカイブとしてのサービスを提供することが可能であると思われる。公的機関が提供する SaaS 型のクラウドサービスとして、国立情報学研究所(NII)が提供する「JAIRO Cloud」がある[5]。NII 内で開発した機関リポジトリソフトウェア WEKO をベースに、大学向けの機関リポジトリ環境を提供しており、2016 年 11 月 30 日時点で 458 機関(公開準備中、申請中の機関を含む)が利用している。JAIRO Cloud は、公的機関が運営・参加し、共通のアプリケーションを利用してデータを共有しているという点で、共同運用型のデジタルアーカイブ・システムに対する先行事例といえる。

また、IaaS では OS レベルからのシステム構築となるが、PaaS においてはアプリケーションシステムの設計・開発がシステム構築の範囲となる。SaaS クラウドサービスの場合は、既存のサービスを組み合わせることにより、アプリケーションサービスを構築することから、システム構築作業はほぼ発生しない。

(3) まとめ

IaaS/PaaS 型のクラウドサービスを利用することで、短期間にかつ安価にインフラを構築できる。また、利用できるサービスは限定されるが、SaaS 型のクラウドサービスによりデジタルアーカイブ・システムを作り上げることができる場合は、さらにコストを抑えることができる可能性がある。システム導入時に数年先の利用状況を予測した上で、サーバの性能を確定し調達する必要がなく、必要な時に必要なだけ増強することができ、導入コストの大幅な軽減が望める。

SaaS 型のクラウドサービスは、特定のアプリケーションソフトウェアによるサービスや機能を提供するものである。デジタルアーカイブ・システムの構築において、SaaS 型のクラウドサービスで利用可能なサ

---

[4] 株式会社矢野経済研究所『2016 クラウドコンピューティング(IaaS/PaaS)市場の実態と展望』(2016 年 7 月 29 日)
https://www.yano.co.jp/press/press.php/001569 (2017 年 1 月 24 日アクセス)

[5] https://community.repo.nii.ac.jp/ (2017 年 1 月 24 日アクセス)

ービスや機能が、各機関の要求に合うか検討が必要である。SaaS 型のクラウドサービスの中にはデジタルアーカイブ・システムのサービスを提供している事例[6]もある。

[6] http://www.amlad.jp/ （2017 年 1 月 24 日アクセス）

2.2.2. サーバ環境

サーバ環境は、システムを動作させるためのハードウェアや OS、ミドルウェア、ストレージ等からなる。サーバ環境で用いられる技術や動向を調査する。

(1) サーバ技術(仮想化技術含む)

コンピュータシステム全般における基盤技術であるサーバ OS、ミドルウェア、サーバソフトウェアと、サーバ機器の性能面での最新動向について調査を行った。また、合わせて、仮想化技術と、仮想化を実現するためのソフトウェアについて、普及度やコストを調査した。

(ア) OS/ミドルウェアの動向、サーバ性能等

サーバ OS では Linux がシェアを伸ばしている。Linux はオープンソースであり初期コストが低く抑えられるうえ、バージョンアップを経て信頼性が向上している。仮想化にも適していることから、導入が進んでいる。

ミドルウェア(Web サーバ、アプリケーションサーバ、データベース)においても、Linux サーバ上で稼働するオープンソースのものが採用される傾向が見られる。これらはベンダーのサポートが受けられないが、初期コストが不要であることに加え、導入事例の増加による情報の蓄積が進んでいることが背景にあると考えられる。デジタルアーカイブ・システムとしてベンダーの提供する有償ソフトウェアも、オープンソースのミドルウェアを組み合わせ、独自のビジネスロジックやユーザーインターフェースを組み込む形で構成されている。

ミドルウェアの中でも、運用監視・システム管理ソフトウェアへの需要が高まっている。クラウド化の進展により、大手企業やデータセンターサービス事業者が、仮想環境のシステム管理最適化やプライベートクラウドにおける運用自動化によるコスト削減に向けて動いていることが背景にあると考えられる。デジタルアーカイブ・システムにおいても、運用監視・システム管理ソフトウェアは、運用コスト削減とサービスの安定稼働に有益な技術として注目される。

サーバ機器の性能という点では、CPU の高クロック化やマルチコア化、メモリの大容量化が進み、全体的に性能向上が見られる。一方で小規模な PC サーバから大規模なブレードサーバまで価格は低下傾向が見られ、価格性能比も向上している。

(イ) 仮想化技術

仮想化技術とは、サーバ機器上で複数の OS を動作させる技術である。従来は 1 台のサーバ機器に 1 つの OS を稼働させており(図 16 仮想化のイメージの「物理マシン」)、複数の OS を稼働させる場合はその数に応じたサーバ機器を用意する必要があった。仮想化技術により、物理的なサーバ台数を最低限に抑えることができるようになった。

仮想化技術はクラウド化においても不可欠な技術であるため、仮想化技術の普及度を推量するために、クラウドインフラストラクチャソフトウェア[7]の市場規模を調査した。2014 年の国内クラウドインフラストラクチャソフトウェア市場は前年比 38.4%増となり、2019 年は 2015 年の 2 倍以上に成長すると予測され

---

[7] クラウド環境(パブリッククラウド、プライベートクラウド両方を対象)を構築、管理するために必要なソフトウェアであり、主にバーチャルマシン/クラウドシステムソフトウェアとシステム管理ソフトウェアで構成される。

ている[8]。仮想化技術も今後、普及が進むと予測される。

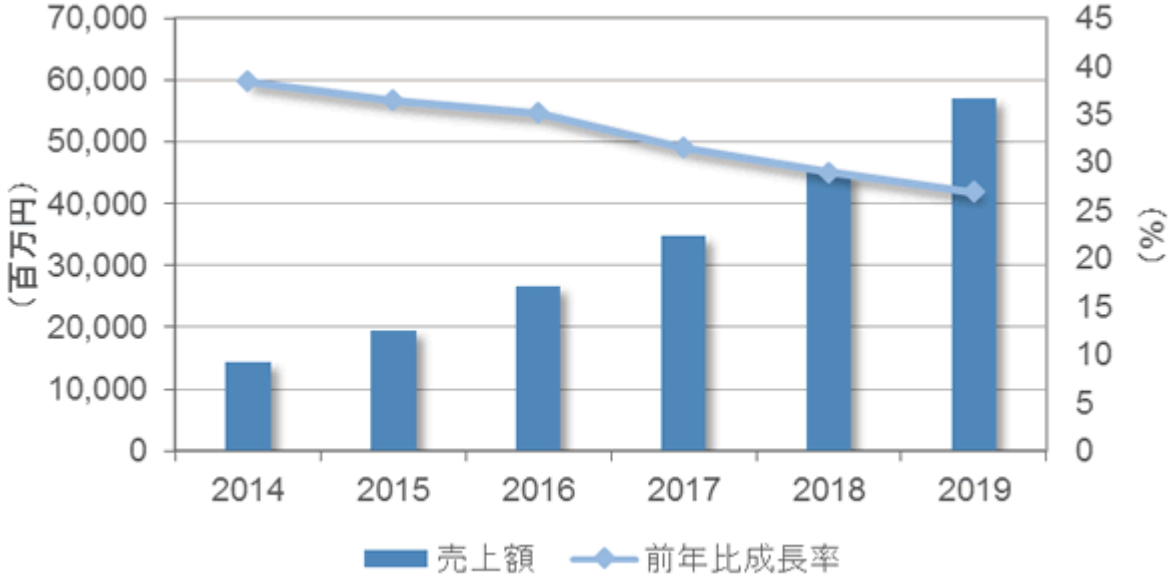


図 15 国内クラウドインフラストラクチャソフトウェア市場 売上額予測：2014 年～2019 年[9]

仮想化技術を実現する仮想化ソフトとしては、有償 OS 付属のソフトウェアとして提供されている Hyper-V がある。また、オープンソースの仮想化ソフトウェアとして、Xen[10]、KVM[11]等がある。これらは図 16 仮想化のイメージの「仮想マシン」に該当する。その他の仮想化ソフトウェアには、コンテナという仕組みにより、仮想マシンとゲスト OS が不要になる Docker[12](図 16 仮想化のイメージの「コンテナ」に該当する)がある。それぞれに固有の特徴や長所を持っている。

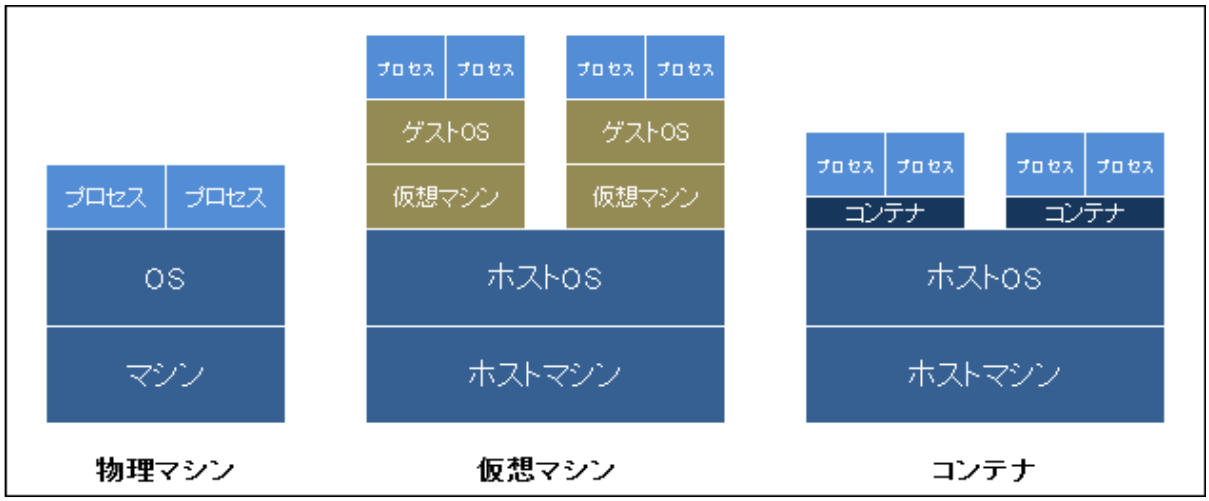


図 16 仮想化のイメージ

---

8 http://www.idcjapan.co.jp/Press/Current/20160118Apr.html（2017 年 1 月 24 日アクセス）
9 注 9 のリンク参照。
10 https://www.xenproject.org/(2017 年 1 月 24 日アクセス)
11 http://www.linux-kvm.org/page/Main_Page（2017 年 1 月 24 日アクセス）
12 https://www.docker.com/（2017 年 1 月 24 日アクセス）

（ウ）まとめ

デジタルアーカイブ・システムのインフラ基盤となるサーバ機器、OS、ミドルウェアは、高機能化や低価格化、オープンソース化が進んでいる。サーバ機器の性能向上に合わせて仮想化技術も発展しており、十分な性能と安定性を持ちながら低コストでデジタルアーカイブ・システムを構築できる環境が整いつつある。

(2)　ストレージ技術

ストレージは、デジタルアーカイブ・システムの各種データを格納する要素である。ストレージの動向、技術について調査する。

（ア）ストレージの動向

ストレージを構成する物理的ディスクとして、従来の HDD に対して、SSD(Solid State Drive)が出荷台数を伸ばしている。SSD は HDD に比べてアクセスが高速であり、消費電力が少ない。ただし、書き込み回数の制約があり、HDD より容量が少なく、容量あたりの単価が高い。

価格については、2000 年以降、面積あたりの記憶機密度が年率 30%から 50%の増加率で向上しており 、これに伴い記憶装置の単価の低下が続いている。市販の HDD の GB あたりの単価に換算すると、1985 年から 30 年間で約 100 万分の 1 まで低下している[13]。SSD でも容量あたりの単価は低下傾向にある。

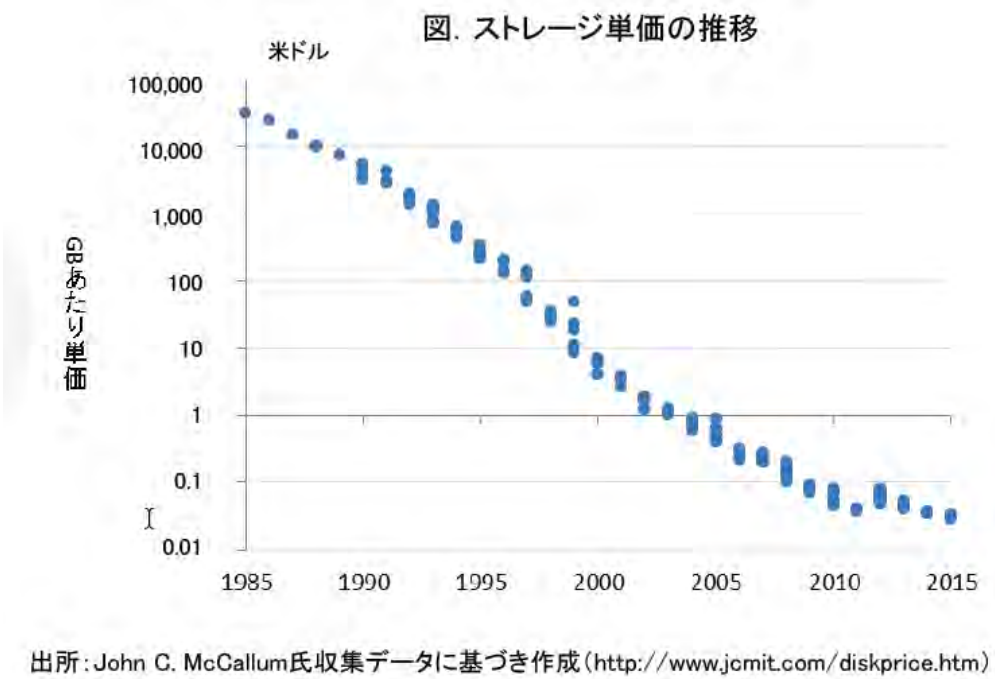


図 17 ストレージ単価の推移

---

[13] 株式会社三菱総合研究所「グローバル ICT 産業の構造変化及び将来展望等に関する調査研究」2015 年 3 月、P9
http://www.soumu.go.jp/johotsusintokei/linkdata/h27_02_houkoku.pdf （2017 年 1 月 24 日アクセス）

（イ） ストレージの技術

システムの信頼性・可用性を確保するため、ストレージには、RAID（Redundant Arrays of Inexpensive Disks）[14]と呼ばれる技術が一般的に用いられる。RAID は複数の物理的なハードディスクドライブ(HDD)を組み合わせ、論理的な HDD を構成する手法であり、信頼性・可用性のレベルに応じて主に RAID0 から RAID6 までの 7 種類またはそれらの組み合わせで構成される。現時点では新たな RAID レベルは出ておらず、技術的には大きな発展は見られない状態である。

技術的な事例として、Apacheトップレベルプロジェクトとしてオープンソースで開発されている「Hadoop」がある[15]。Hadoop は Hadoop Distributed File System (HDFS)と呼ばれるファイルシステムを用いて、複数のノードに分散してファイルを配置することで、ペタバイト級の大規模データを効率的に処理できる仕組みとなっている。また、異なるノードに複製を置くことで信頼性を高めている。現在はビッグデータの分散処理等に用いられる技術だが、将来的にデジタルアーカイブ・システムが大量データを処理する際に活用できる可能性がある。

デジタルアーカイブ・システムにおいてストレージを選定するにあたり、信頼性確保のため RAID 構成を組めることが必要である。HDD と SSD は性能・コストを考慮したうえで適切なものを選択すればよいと考えられる。

（ウ） クラウドストレージ

「2.2.1 クラウド技術」で調査したクラウドサービスには、クラウドストレージのサービスが提供されており、使用量に応じた従量課金や、一定量までの月額料金で利用できる。但し、自館外にデータを配置する場合、データの漏洩や消失を防ぐための対策が必要であるため、自組織のセキュリティポリシーに従って、配置するデータの種類や内容を検討する必要がある。また、重要なデータのバックアップは各機関で取得するといった対策が必要となる。

（エ） まとめ

ストレージ技術において、その基礎となる技術は普及してから相当の年月が立っており、安定した技術として提供されている。ストレージの容量あたりの単価は減少傾向にあり、デジタルアーカイブ・システムで大容量のデータを扱うためのコスト面でのハードルは低くなりつつある。また、クラウドストレージを使うことで、必要な容量のみをコストを抑えて用意することが可能である。

---

[14] David A Patterson, et al. A Case for Redundant Arrays of Inexpensive Disks (RAID), University of California Berkeley, 1988

[15] http://hadoop.apache.org/ （2017 年 1 月 24 日アクセス）

2.2.3. ネットワーク環境

本項では、サーバとインターネットを接続するための通信回線に関する技術的な最新動向、普及度やコストについて調査する。

(1) 通信回線

現在、インターネット接続回線を提供するネットワークプロバイダとして多くの企業が参入しており、様々なサービスが提供されている。これらのサービスで提供されている回線で主に使用されている技術には、既存の電話回線を利用する DSL[16]と、光ファイバーを伝送路に使用する光回線がある。DSL のうち日本国内では ADSL(非対称デジタル加入者線)が 2001 年以降一般家庭にも普及したが、より高速な光回線が普及したことで加入者数は減少の一途をたどっている。ケーブルテレビ(CATV)の回線を用いたインターネット接続も加入者数が増えつつある。

図 18 固定ブロードバンドサービスの加入者数推移[17]

最大通信速度は ADSL が 50.2Mbps、光回線が 1Gbps であり、光回線が優位である。代表的な事業者の価格は ADSL が月額 5 千円程度[18]、光回線が月額 4 万円程度[19]であり、価格面では ADSL が低く抑えられる。ただし、ADSL は加入者数の減少に伴い、新規加入を停止している事業者があることに注意が必要である。

各機関がすでにインターネット回線を用意している場合はその回線を使うことが合理的な選択である。ただし、ファイルサイズが大きいコンテンツを配信する場合は、1Gbps 程度の帯域を持った回線を用意す

---

16 デジタル加入者線。
17 http://ictr.co.jp/report/20160712.html (2017 年 1 月 24 日アクセス)
18 フレッツ・ADSL ADSL 専用型 モア III(47M タイプ)
https://flets.com/adsl/fee.html (2017 年 1 月 24 日アクセス)
19 フレッツ 光ネクスト ビジネスタイプ
https://flets-w.com/next/ryoukin/business.html (2017 年 1 月 24 日アクセス)

ることが望ましい。

(2)　暗号化技術

デジタルアーカイブ・システムでは、職員が利用する業務側の機能では非公開のデータを取り扱うケースもある。インターネットを経由して非公開データを扱う場合には、通信の暗号化は重要な要素である。本項では、インターネット上での暗号化技術について調査を行う。

インターネット、特に Web 上では HTTP(Hypertext Transfer Protocol)に SSL/TLS(Secure Sockets Layer/Transport Layer Security)プロトコル[20]による暗号化を加えた HTTPS(Hypertext Transfer Protocol Secure)が使用されている。ただし、SSL/TLS には脆弱性が多く見つかっており、TLS1.2 への移行が Web サイト、Web ブラウザの両方で進められている。

Web 上での HTTPS リクエストの割合は 3 割程度であるが、米連邦政府は政府機関 Web サイトを 2016 年 12 月 31 日までに HTTPS 化(入力フォーム等による利用者情報の送信を含まない Web ページも含めて、すべてを SSL/TLS による暗号化の対象とすること)を義務化しており[21]、主要な OS・ブラウザベンダーでも HTTPS 化の動きがあるため、HTTPS への対応は今後促進されていくと考えられる。

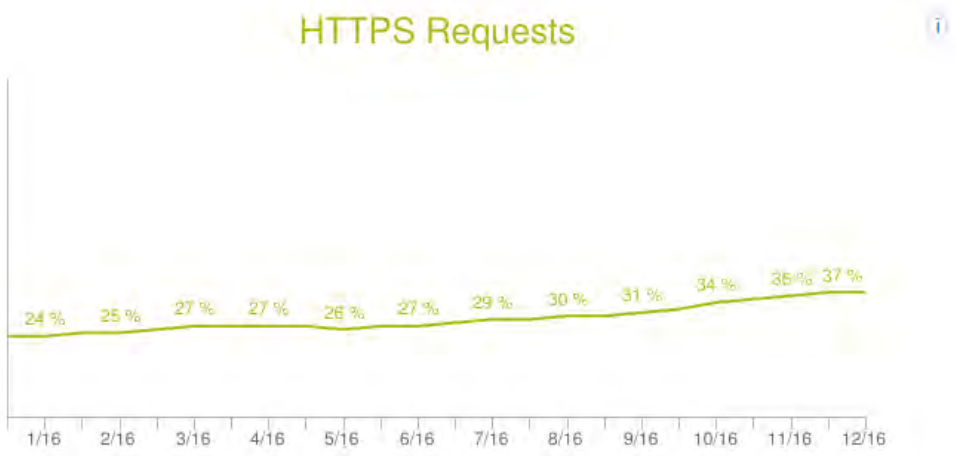


図 19 HTTPS リクエストの割合の推移[22]

また、インターネット上におけるデータ交換による電子商取引の拡大に伴い、通信上のなりすまし、改ざん、盗聴を防ぐために、信頼された認証局により発行される公開鍵証明書(SSL サーバ証明書)が広く用いられている。公開鍵証明書は、情報の暗号化に用いる公開鍵というデータと所有者との関係を第三者機関である認証局が認証するための情報である。認証にはドメイン認証、企業認証、EV 認証の 3 つのレベルがあり、認証局によって年額数千円から 20 万円近いものまでさまざまなものがある。

---

[20] SSL を元に TSL が制定された経緯があり、慣習的にこのように表記される。
[21] https://cio.gov/https-everywhere-for-government/ (2017 年 1 月 24 日アクセス)
[22] HTTP Archive より(2015 年 12 月 15 日から 2016 年 12 月 16 日までの情報を取得)
http://httparchive.org/trends.php?s=All&minlabel=Dec+15+2015&maxlabel=Dec+2+2016 (2017 年 1 月 24 日アクセス)

(3)　HTTP/2

HTTP/2 (Hypertext Transfer Protocol version 2) は、Web 上で普及している HTTP の新バージョンである。HTTP/2 は 2015 年 2 月 17 日に正式な仕様として承認され、2015 年 5 月に RFC 7540 として文書化された[23]。

HTTP/2 は従来の HTTP との互換性を維持しながら、Web の高速化・効率化を図ることを目的としている。HTTP/1.1 では 1 つのリクエストが完了するまで原則として次のリクエストは送信できず、同時接続数にも制約があるため多くの画像コンテンツ等を含む Web サイトではリクエスト数が増え、読み込みの効率が悪いという問題があった。HTTP/2 では、同一ホストへの複数のリクエストをストリームという単位で多重化し、フロー制御を加えることで複数のリクエストを同時に処理できるようになり、Web サイト読み込みの高速化を実現できるようになった。

現在の主要な Web ブラウザ、Web サーバの最新バージョンが HTTP/2 に対応している。ただし、Web サーバ側で追加モジュールや設定が必要となる。HTTP/2 は TLS による暗号化は必須としていないが、主要な Web ブラウザには HTTP/2 において HTTPS での接続のみしか許可しないという方針を表明しているものがある。

デジタルアーカイブ・システムでは、複数のデジタルコンテンツのサムネイルを表示するといったケースが想定され、HTTP/2 を採用することで利用者が高速に閲覧できると考えられる。

(4)　まとめ

デジタルアーカイブ・システムにおいて、各機関がすでにインターネット回線を用意している場合はその回線を使うことが合理的な選択である。ただし、ファイルサイズが大きいコンテンツを配信する場合は、1Gbps 程度の帯域を持った回線を用意することが望ましい。公開するコンテンツ、想定するアクセス数等に応じた適切な回線を用意することが必要である。

HTTP の最新版である HTTP/2 への対応は、閲覧速度向上により利用者のストレス軽減という効果が期待できる。

暗号化については、将来的に HTTP/2 への移行が進んだ場合に HTTPS が必要となること、利用者のセキュリティへ配慮することを考えると、HTTPS をサポートすることが望ましいと考えられる。

---

[23] https://tools.ietf.org/html/rfc7540 （2017 年 1 月 24 日アクセス）

2.2.4. バックアップ関連技術

デジタルアーカイブ・システムに関連するデータは、表 8 のように分類できる。

表 8 各機関における公文書等のデータの分類

データ区分	対象データ
保存用データ	電子公文書のオリジナル、デジタル化した画像等のデジタルコンテンツデータ
業務用データ	非公開の情報を含む目録データ、非公開情報を含む画像等のデジタルコンテンツデータ、業務上必要な管理データ等
公開用データ	一般公開が可能な目録データ、デジタルコンテンツデータ等
システムデータ	デジタルアーカイブ・システムのシステム(OS、ミドルウェア、アプリケーション)に関するデータ

バックアップ技術の基礎となるバックアップ方式と、使用されるメディアの最新動向に関して調査を行った。また、東日本大震災を契機に、BCP として遠隔地にバックアップを保管することの重要性が見直されている。その状況を踏まえて、バックアップ拠点に関して調査する。

(1) バックアップ技術

システムの運用により密接なバックアップ技術に、ホットバックアップ、コールドバックアップの 2 つがある。

ホットバックアップとは、システムを停止することなくバックアップを行う技術である。ホットバックアップでは、専用ソフトウェアや専用ハードウェアが必要となり、バックアップ取得方法が複雑化するため、コストが高くなる。

コールドバックアップとは、システムを停止してバックアップを行う技術である。バックアップを取得中はシステムが使用できないが、OS レベルでのファイルコピー等の方法でバックアップ取得を行うことができる。バックアップ取得方法が簡素化され、コストを低く抑えることができる。

これらのバックアップ技術を表 9 で比較する。

表 9 バックアップ技術の違いによるコストの比較

	システム停止	コスト
ホットバックアップ	不要	高
コールドバックアップ	必要	低

システムの運用において、システムの稼働目的から、1 日のうち時間帯を区切って稼働するシステムや、年間を通して極力停止しないことを求められるシステム等、多様な運用スケジュールが考えられる。このような運用の実態に合わせ、採用するバックアップ技術を検討することにより、バックアップ計画の最適化や必要となるコストの合理化を図ることが可能である。

(2) データのバックアップ方式

バックアップ方式は、完全バックアップ、増分バックアップ、差分バックアップの 3 つに分類される(図

20)。

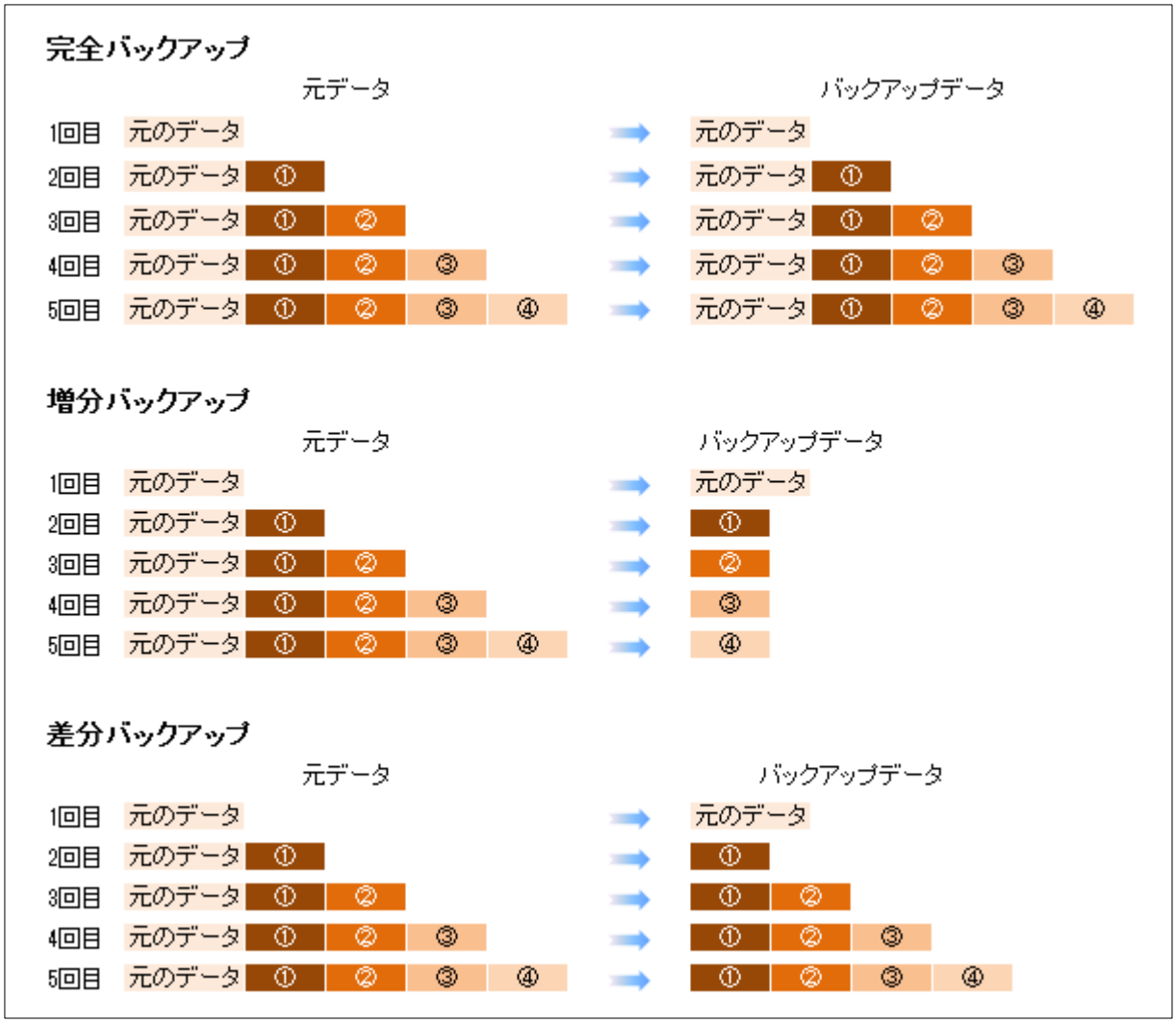


図 20 バックアップ方式のイメージ

完全バックアップでは、毎回全てのデータがバックアップ取得範囲となるため、バックアップデータを格納するために大容量のメディアが複数必要となる。

増分バックアップの場合には、最初のみ全てのデータがバックアップ取得範囲となるが、2 回目以降は前回取得したバックアップから増えたデータのみがバックアップ取得範囲となるため、完全バックアップに比べ格納メディアの数は抑えられる。

差分バックアップでは、増分バックアップと同様に最初のみ全てのデータがバックアップ取得範囲となるが、2 回目以降は最初から増えた分だけのデータがバックアップ取得範囲となり、格納メディアの数は完全バックアップよりは抑えられるが増分バックアップよりは多くなる。

(3) バックアップメディア

バックアップで用いられるメディアは、従来から利用されている磁気テープが主流であるが、入出力速度が速く短時間で復旧を行えるハードディスクや、長期間の保存が可能で低コストの光学ディスクの採用

も増えつつある。光学ディスクでは、耐久性に優れている上、マイグレーション[24]の頻度を抑えることが可能という性質がある。SSD は、ハードディスクより入出力速度が速いが、自然放電によりデータを消失する可能性がある。半導体メディア(SD カード)は、ハードディスクより入出力速度が速いが、大容量のものがなく、SSD と同様に、自然放電によりデータを消失する可能性がある。

バックアップ容量あたりのコストは、磁気テープ、ハードディスク、光学ディスクの順で高くなる傾向がある。バックアップメディアの比較を表 10 にまとめる。

表 10 バックアップメディアの比較

	入出力速度	耐久性	マイグレーション	コスト
磁気テープ	遅	低	要	低
ハードディスク	速	中	要	中
光学ディスク	中	高	不要	高
SSD	最速	低	要	高
半導体メディア (SD カード)	最速	低	要	高

磁気テープは入出力速度が遅いが容量あたりの単価が低いため、大容量のデータをバックアップする用途に適している。ハードディスクは入出力速度が高速なため、短時間でデータをバックアップする用途に適している。光学ディスクはマイグレーションが不要という性質から、コールドデータ[25]が求める長期保存に適している。

一方、SSD や半導体メディア(SD カード)は入出力速度が速いため、作業途中のデータ等一時バックアップとしての用途に適している。長期間通電しないと、自然放電によりデータを消失するという特性を考慮すると長期保存には適さない。

(4) バックアップ拠点

バックアップ拠点は、以下の 3 つに分類できる。

(ア) オンサイトバックアップ

オンサイトバックアップは、バックアップ用の機器を導入(サーバ内蔵または外付けがある)し、随時もしくは定期的にバックアップを取得し自組織内で保管する方式である。サーバと同じ場所にバックアップがあるため、データの漏洩や紛失のリスクがなく、リストアが短時間で行える。ただし、大規模災害が発生した場合にバックアップデータも消失してしまうリスクがある。

(イ) 遠隔地バックアップ

遠隔地バックアップは、バックアップデータを自組織外の場所に保管する方式である。データの受け渡し方法により、オンライン型と搬送型に二分される。オンライン型では、国内外を問わず複数拠点にある

---

24 媒体の経年変化対策として、数年ごとに異なる媒体に複製する作業。
25 時間が経過すると使う頻度は低くなるが、長期保存が必要となるデータ。

データセンターにネットワーク経由でデータをバックアップし保管する。搬送型では、物理媒体に保管されたデータを専用の搬送容器で搬送し保管する。

オンライン型は搬送型よりもバックアップからの復旧が迅速に行えるが、大規模災害が発生し、ネットワーク回線の断線が発生した場合は、復旧までに時間を要する可能性もある。また、ネットワークセキュリティや保管先の信頼性も考慮する必要がある。

搬送型は低コストで実現できるが、復旧には再度の搬送が必要となり時間がかかること、専用搬送容器を使うとしても紛失や盗難のリスクがある。

（ウ）クラウドバックアップ

クラウドバックアップは、バックアップデータをクラウドサービス上に保存する方式である。前項の遠隔地バックアップのうちオンライン型と同様であるが、サービス事業者によりバックアップデータの容量、暗号化といったサービス内容に差がある。ネットワークセキュリティや保管先の信頼性を考慮すべき点はオンライン型と同様である。

(5)　まとめ

他のシステム同様、デジタルアーカイブ・システムにおいても、必要となるコストを踏まえた効率的な運用を実現するために、各機関の保有するデータ量と、将来的なデータ量の増加予測に応じて、適切なバックアップ方式やメディアを検討し導入することが推奨される。

また、遠隔地バックアップは各機関が BCP 策定時にコストや必要性を踏まえて検討すべきある。

表 11 にバックアップの推奨例を挙げる。適していると考えられるものを○、適さないものを×、各機関の事情により適すると考えられるものを△とする。

表 11 バックアップの推奨例

		保存用データ	業務用データ	公開用データ	システムデータ
バックアップ技術	ホット	×	×	○	○
	コールド	○	○	△	△
バックアップ方式	完全	○	○	△	△
	差分	△	△	○	○
	増分	△	△	○	○
バックアップメディア	磁気テープ	△	○	○	○
	ハードディスク	△	○	○	○
	光学ディスク	○	○	△	×
	SSD	×	×	△	△
	半導体メディア（SD カード）	×	×	×	△
バックアップ拠点	オンサイト	△	△	△	△
	遠隔地	○	○	△	△
	クラウド	×	×	○	○

バックアップ技術について、公開中のデジタルアーカイブ・システムに影響を及ぼさない保存用データ、

業務用データについては、コールドバックアップが適している。逆に公開中のデジタルアーカイブ・システムに影響を及ぼす公開用データ及びシステムデータについては、システム停止を伴わないホットバックアップが望ましい。しかしながら、利用の少ない夜間等にシステムを停止させ、コールドバックアップを取得することも選択肢として考えられる。

バックアップ方式について、資料のデジタル化に伴い徐々に増える保存用データ、業務用データについては、長期的に保存することを考慮し、完全バックアップが適している。一方、公開用データ、システムデータについては、公開データの追加及びシステム更新の頻度に応じて、差分バックアップもしくは増分バックアップを選択することが考えられる。

バックアップメディアについては、各データのサイズ及びデジタルアーカイブ・システム復旧時の読出し速度を考慮する必要がある。保存用データは大容量かつ長期的に保持することから光学ディスクが適している。業務用データは保存用データよりも保持期間が短く済むため、磁気テープ、ハードディスク、光学ディスクが適している。公開用データは磁気テープ、ハードディスクが適しているが、容量とかけられるコストによっては光学ディスク、SSD も選択肢となりうる。システムデータは、コスト面から磁気テープ、ハードディスクが適しているが、他のデータと異なり容量は小さくなるため、SSD、半導体メディアについても選択肢となりうる。

バックアップ拠点については、各機関の BCP に応じて検討する必要がある。大災害により建屋が倒壊した場合を想定すると、いずれのデータでもオンサイトバックアップにはリスクがある。保存用データ、業務用データについては長期的に保持する必要があり、データの重要性も高いことから遠隔地バックアップが望ましい。一方、公開用データ及びシステムデータについては、遠隔地バックアップも選択肢となりうるが、この場合システム復旧までに時間がかかる。迅速なシステム復旧という観点では、クラウドバックアップが適している。なお、大災害によって、ネットワーク回線の断線が発生した場合、クラウドバックアップからの復旧に時間を要することが想定されるが、その際は、デジタルアーカイブ・システムを公開するためのネットワーク回線の断線も想定される。よって、いずれにしてもサービス再開に時間を要するものと考えられ、クラウドバックアップであることに大きな支障はないと考えられる。

2.2.5. ネットワークセキュリティ技術

デジタルアーカイブ・システムのみならず、インターネット上で公開されているシステムはサイバー攻撃の対象となりうる。サイバー攻撃は安定的なサービスの提供、データ保全の大きな脅威であり、セキュリティ対策は非常に重要である。本項では、サイバー攻撃手法及びその対策について最新動向を調査する。

(1) 攻撃手法

独立行政法人 情報処理推進機構(IPA)の発表した 2016 年の情報セキュリティ 10 大脅威[26]では、個人・組織を標的として情報の不正利用や流出を狙うものが上位に位置付けられているが、組織を標的にしたものではサービス妨害攻撃によるサービスの停止、Web サイトの改ざん、脆弱性の悪用といったサービス継続に影響を及ぼす脅威が、より重大な脅威とされている。

タイトル(個人)	順位	タイトル(組織)
インターネットバンキングやクレジットカード情報の不正利用 総合:1位	1	標的型攻撃による情報流出 総合:2位
ランサムウェアを使った詐欺・恐喝 総合:3位	2	内部不正による情報漏えいとそれに伴う業務停止 総合:8位
審査をすり抜け公式マーケットに紛れ込んだスマートフォンアプリ 総合:7位	3	ウェブサービスからの個人情報の窃取 総合:4位
巧妙・悪質化するワンクリック請求 総合:9位	4	サービス妨害攻撃によるサービスの停止 総合:11位
ウェブサービスへの不正ログイン 総合:5位	5	ウェブサイトの改ざん 総合:6位
匿名によるネット上の誹謗・中傷 総合:13位	6	脆弱性対策情報の公開に伴い公知となる脆弱性の悪用増加 総合:10位
ウェブサービスからの個人情報の窃取 総合:4位	7	ランサムウェアを使った詐欺・恐喝 総合:3位
情報モラル不足に伴う犯罪の低年齢化 総合:18位	8	インターネットバンキングやクレジットカード情報の不正利用 総合:1位
職業倫理欠如による不適切な情報公開 総合:14位	9	ウェブサービスへの不正ログイン 総合:5位
インターネットの広告機能を悪用した攻撃 総合:12位	10	過失による情報漏えい 総合:16位

図 21 情報セキュリティ 10 大脅威 2016 個人・組織別順位[27]

サービス妨害攻撃の攻撃手法としては、ボットネットと呼ばれるプログラムを用いて、標的となる組織のルータ、DNS、サーバへ負荷をかける方法が用いられている。脆弱性を悪用して乗っ取られた外部組織のサーバが攻撃の踏み台として使用されるケースもある。

Web サイトの改ざんでは、OS・ミドルウェア、アプリケーションの脆弱性を突いてサーバへ侵入したり、

---

[26] 「情報セキュリティ 10 大脅威 2016」独立行政法人 情報処理推進機構 セキュリティセンター、2016 年 3 月 https://www.ipa.go.jp/files/000051691.pdf (2017 年 1 月 24 日アクセス)

[27] 前掲、P16

外部からのアップロードが可能になっているといった設定の不備を突いて悪意のあるプログラムを埋め込んだりといった攻撃手法が取られている。

(2) 対策

内部不正や過失による情報漏えいといった脅威は、組織の構成員に対する情報セキュリティ教育や体制面での対応といった対策が重要であるが、脆弱性の悪用といった脅威は、定期的に OS・ミドルウェアを更新することにより脅威を軽減することができる。また、サービス妨害攻撃や、Web サイトの改ざん等の脅威に対して有効と思われる、ネットワークセキュリティ技術を用いた対策について、特徴とコストを表 12 に記載する。

表 12 セキュリティ脅威に対する対策

対策	説明	コスト
不正侵入検知/防御(IDS/IPS)装置	ネットワーク上の不正なトラフィックを検知し、遮断するための機器またはソフトウェア。検知・防御範囲が広い。 IDS は侵入の検知のみ行い、IPS は侵入を防ぐ機能を持つ。	高 専用の機器、ソフトウェアの導入、設定が必要。
改ざん検知ソフトウェア	Web コンテンツの改ざんを検知するソフトウェア。 Web コンテンツのファイル更新を監視するものから、攻撃パターンとの一致を検出するもの、Web ブラウザ上での動作から改ざんを検知するものまで多くの方式がある。	中 ソフトウェアにより検知レベルは異なるが、導入、設定が必要。
ファイアウォール装置	ネットワーク上の通信パケットの送信元・送信先アドレス、ポート番号から通過の可否を判断するための機器またはソフトウェア。	高 専用の機器、ソフトウェアの導入、設定が必要。
Web Application Firewall(WAF)	Web サーバと Web ブラウザ間の HTTP/HTTPS 通信の内容を検査し、攻撃を検出するためのソフトウェア。 ファイアウォール装置がネットワークレベルでの防御を行うのに対して、WAF はアプリケーションレベルでの防御を行う。	中～低 ソフトウェアの導入、設定が必要。
ウイルス対策ソフト	悪意を持ったウイルスファイルを検知し、駆除することでコンピュータ上の脅威を軽減させるためのソフトウェア。 新種のウイルスへ対応するためにはウイルスパターンファイルを定期的に更新する必要がある。	中～低 ソフトウェアの導入、設定が必要。

WAF については、Web サーバ「Apache」のモジュールであるオープンソースの「ModSecurity」がある[28]。IPA では ModSecurity の導入・運用事例を紹介しており、セキュリティレベルの向上に有益な情報を提供している[29]。

また、IPA が公開している「安全なウェブサイトの作り方」[30]では、アプリケーションを実装する上での具体的な対策が示されており、それらの対策の実装状況を確認するためのツールとして、同資料の別冊「セキュリティ実装 チェックリスト」が公表されている。ほかに内閣サイバーセキュリティセンターが公表している「情報システムに係る政府調達におけるセキュリティ要件策定マニュアル」やセキュリティベンダー提供する情報等もデジタルアーカイブ・システムを新規開発またはカスタマイズする際のセキュリティ対策の指針として有用であると考えられる。

(3) まとめ

サイバー攻撃の手段は多様化・複雑化しており、デジタルアーカイブ・システムにおいても安定的なサービスの提供、データ保全のために十分なセキュリティ対策を講じる必要がある。システム上の対策として様々な選択肢があるが、セキュリティレベルに応じてコストが高くなるため、各機関のセキュリティ対策方針に応じて選択する必要がある。最低限の対策として、ウイルス対策ソフトと WAF の導入が推奨される。また、デジタルアーカイブ・システムの構築事業者に対して、IPA 等の公表資料で示されるセキュリティ対策への準拠を求めることも有効な手段であると考えられる。

---

28 http://www.modsecurity.org/ （2017 年 1 月 24 日アクセス）

29 『Web Application Firewall(WAF)読本」』独立行政法人 情報処理推進機構 セキュリティセンター、2011 年 2 月
http://www.ipa.go.jp/files/000017312.pdf （2017 年 1 月 24 日アクセス）

30 https://www.ipa.go.jp/security/vuln/websecurity.html （2017 年 1 月 24 日アクセス）

2.2.6. 目録関連技術

アーカイブズにおける記述標準としては ISAD(G)（General International Standard Archival Description）とその XML 表現である EAD(Encoded Archival Description)がある。本項では、EAD 及びその他のデジタルアーカイブ・システムに関連する目録関連技術の最新動向について調査する。

（1） EAD

本項では、符号化記録史料記述である EAD の最新版である EAD3 について、前バージョンの EAD2002 との差異を調査し、デジタルアーカイブ・システムにおける広範な目録情報交換の可能性について検討する。

（ア）記録史料記述の標準化動向

記録史料記述における国レベルでの標準には、DACS（Describing Archives: A Content Standard、米国）、RAD（Rule for Archival Description、カナダ）、MAD（Manual of Archival Description、英国）等がある。国際標準の記述方式としては、1992～96 年に国際公文書館会議（ICA：International Council on Archives）が制定した ISAD(G)がある。

ISAD(G)は記述に関する考え方で、これをホームページ等の作成に用いるマークアップ言語の XML（Extensible Markup Language）形式で表現した符号化記録史料記述が EAD（Encoded Archival Description）である。EAD は、目録の符号化における事実上の国際標準であるとともに、目録データを階層構造で表現することが可能である。

EAD は 2002 年に 1.0 から EAD2002 へ改版され、2015 年 8 月に EAD 3(1.0)が正式に公開された[31]。

（イ）EAD2002 と EAD3 の比較

EAD2002 から EAD3 への改訂にあたっては、以下の四点が重視されている。

1. EAD 利用における概念的・意味論的な一貫性の確保
2. EAD でエンコードされた情報を、他のプロトコルにおいてシームレスかつ効率的にデータ結合・交換を行える探索メカニズム
3. 多言語環境において作成された情報を表現する手段・機能の追加
4. 現行 EAD ユーザーに対する新バージョンによる影響への配慮

EAD3 における主な変更点は以下の通りである[32]。

- <ead>要素の子要素から<eadheader>が廃止され、管理情報を記述するための代替手段として<control>要素が追加された。
- EAD3 と他システムとの情報交換を容易にするため、<did>要素の子要素が広範囲に渡り更新された。たとえば、<origination>、<langmaterila>等の要素で、曖昧さを排除するため、テキストとタグの混在が許可されなくなった。
- 空要素を除くすべての要素に、多言語記述をサポートするための@lang 属性及び@script 属性

---

[31] https://www.loc.gov/ead/ead3available.html （2017 年 1 月 24 日アクセス）
[32] https://www.loc.gov/ead/EAD3taglib/index.html （2017 年 1 月 24 日アクセス）

が追加できるようになった。

- 概念的・意味論的な一貫性を確保するための多数の変更が加えられた。たとえば、<arrangement>要素が<scopecontent>要素の子要素ではなくなり、<unitdate>要素が<unittitle>要素の子要素ではなくなった、等。

（ウ）EAD3 の導入事例

EAD3 は 2015 年に公開されたが、本調査の時点（2016 年 12 月）において、EAD3 に対応したシステムの実装は世界的にも確認できなかった。しかし、欧州の文書館ポータルサイトである Archives Portal Europe が EAD3 の実装に着手することを公表している[33]。

(2) Dublin Core

デジタルアーカイブ・システムでは、利用者が探す資料に容易に辿り着くためには各機関の持つ目録情報のみならず、他の機関の目録情報を同時かつ横断的に検索できる仕組みは有用である。横断検索では意味の共通性において、共通の目録項目を用いて検索することが重要である。本項では、メタデータの統一的な語彙である Dublin Core に関して、技術的な最新動向、事例を調査し、デジタルアーカイブ・システムにおける横断検索への適用可能性について検討する。

（ア）Dublin Core の現状

Dublin Core は基本的な 15 エレメント(Dublin Core Metadata Element Set, DCMES[34])を定義することによってインターネット上の多様な情報資源に対応し、情報資源発見の効率化を図ることを目的として定義された、コアとなるメタデータ集合である。Dublin Core が担う最も重要な役割は相互運用性（Interoperability）への寄与であり、Web コンテンツ、既存の各種メタデータなど多様な情報資源を Dublin Core 介して相互に結ぶことで、情報資源の発見が容易になる。2003 年には ISO 15836 及び NISO Z39.85 によって規格化された。

Dublin Core の基本 15 エレメントは、様々な分野のメタデータから共通の項目を抽出しているため、扱いが容易で簡潔である。しかし、データ項目の数が多くない分、表現可能な情報の粒度が粗いことから、基本 15 エレメントを拡張し、定義域と値域を加えた DCMI Metadata Terms(DC タームス)[35]が 2008 年に公開されている。

（イ）アプリケーションプロファイル

アプリケーションプロファイルとは、複数の語彙定義の要素集合から、実際の応用（アプリケーション）ごとに必要な要素を選ぶとともに構造制約を定義したものである[36]。アプリケーションプロファイルの基本的な考え方は、すでに流通している語彙を利用することで語彙定義にかかるコストを下げ、相互運用性

---

33 http://wiki.archivesportaleurope.net/index.php/EAD3 （2017 年 1 月 24 日アクセス）
34 http://dublincore.org/documents/dces/ （2017 年 1 月 24 日アクセス）
35 http://dublincore.org/documents/dcmi-terms/ （2017 年 1 月 24 日アクセス）
36 『平成 22 年度 新 ICT 利活用サービス創出支援事業（電子出版の環境整備）メタデータ情報基盤構築事業報告書』一般社団法人 メタデータ基盤協議会、2011 年 3 月 31 日、P102
http://www.mi3.or.jp/item/metaproj2010.pdf （2017 年 1 月 24 日アクセス）

を高めることである。

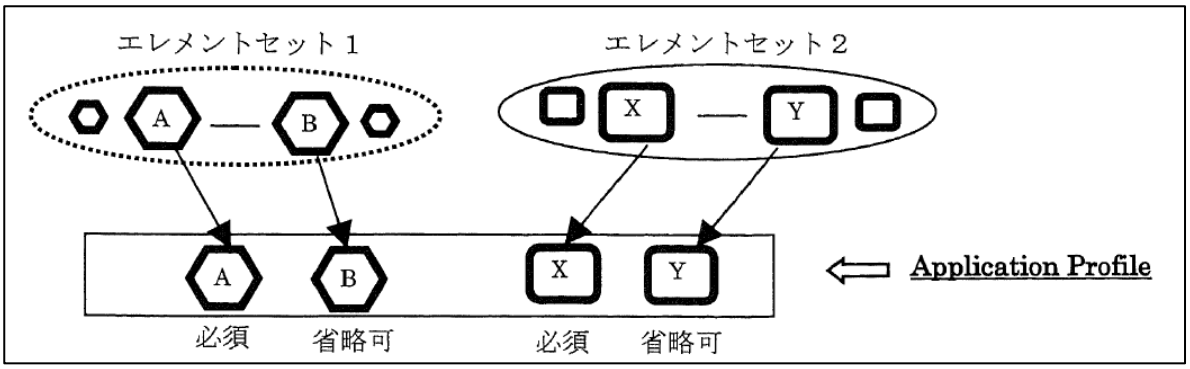


図 22 アプリケーションプロファイルの概念図[37]

DCMES と DC タームスによりメタデータの標準化が図られたが、特定のコミュニティやアプリケーションにおけるメタデータのニーズは、それぞれ大きく異なり、ニーズに応じたメタデータフォーマットが増えることとなった。その解決策として考えられたのがアプリケーションプロファイルであり、Dublin Core におけるアプリケーションプロファイルの定義を示す枠組みとして「シンガポールフレームワーク[38]」が提唱されている。

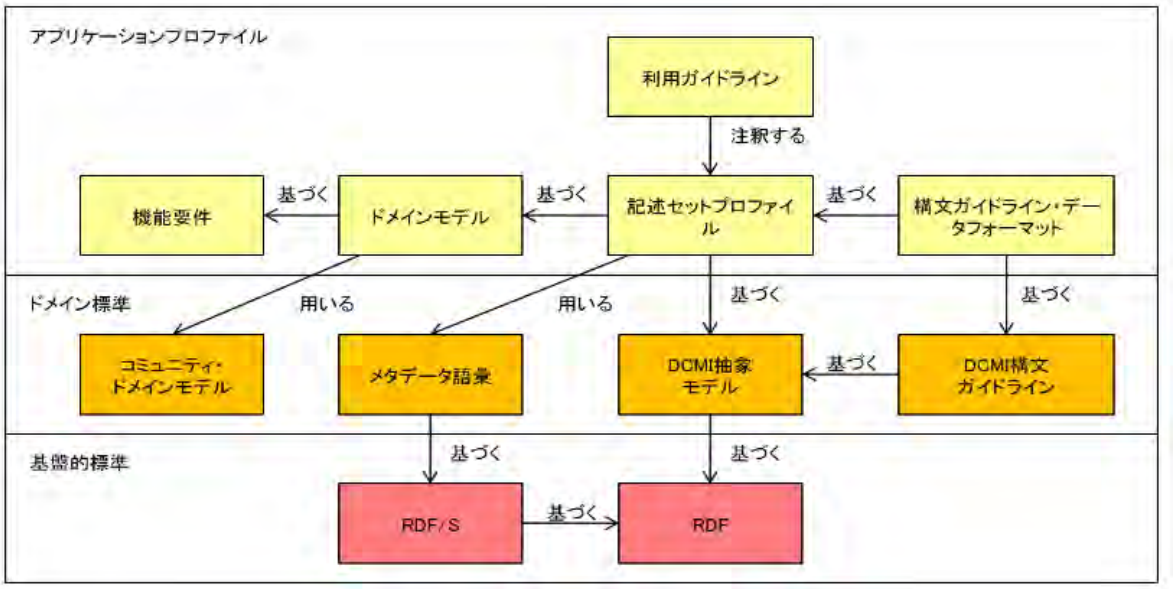


図 23 シンガポールフレームワークにおけるアプリケーションプロファイルの構成図[39]

---

37 杉本重雄「Dublin Core の現在」 ディジタル図書館編集委員会『ディジタル図書館』36 号、2009 年 3 月、P32-45
つくばリポジトリ
https://tsukuba.repo.nii.ac.jp/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=22284&item_no=1&page_id=13&block_id=83 （2017 年 1 月 24 日アクセス）

38 個々のメタデータ作成・提供機関がアプリケーションプロファイルを策定する際に依拠すべき枠組みとして、ダブリンコアが提唱しているもの。

39 「アプリケーションプロファイルのためのシンガポールフレームワーク」 国立国会図書館電子情報部電子情報流通課 訳、2008 年 1 月
http://ndl.go.jp/jp/aboutus/standards/translation/singapore.htm （2017 年 1 月 24 日アクセス）

（ウ）アプリケーションプロファイルを用いた事例

国内では、国立国会図書館ダブリンコアメタデータ記述（DC-NDL）が、アプリケーションプロファイルの枠組みを採用し、DC タームスを用いてメタデータを記述している[40]。DC-NDL では、図書以外の情報資源全般に係るメタデータを表現するために、DCMES、DC タームスの他に FOAF(Friend of A Friend)[41]、RDF(Resource Description Framework) Schema[42]、OWL(Web Ontology Language)[43]といった主要なメタデータ標準語彙を積極的に採用し、国立国会図書館の独自語彙（日本語独自の「よみ」やサービス要件に関するもの等)を含めて「NDL Metadata Terms」として公開している[44]。DC-NDL は図書や雑誌記事等のメタデータを主眼としたものであるが、デジタルアーカイブ・システムにおいてもアプリケーションプロファイルの考え方や DCMES・DC タームスと目録情報のマッピングを考慮する上で非常に有益な事例であるといえる。

海外では、米国デジタル公共図書館(Digital Public Library of America、DPLA)において DPLA Metadata Application Profile(MAP)と呼ばれるアプリケーションプロファイルが公開されている[45]。DPLA MAP は図書館、文書館、博物館のデータモデルを適合させるものであり、Europeana のデータモデルでの経験や DPLA コミュニティからのフィードバックを踏まえて作成されている。

(3) パーマリンク

（ア）パーマリンクとは

パーマリンク(パーマネントリンク(Permanent Link)の略)とは、Web サイトの各ページ、コンテンツに対して永続的に与えられる URL(URI)を指す。パーマリンクを用いることで、Web サイトで提供するコンテンツへのアクセスが長期的に保証される。記録史料を次世代まで活用できるように蓄積・保存・公開していく役割を持つ機関にとって、デジタルアーカイブ・システムのコンテンツが長期的に安定して利用可能であることは、特に重要な機能といえる。

また、次項「2.2.7 目録データの提供」で触れる RDF や LOD といった技術、「2.2.9 デジタルコンテンツ配信技術」で触れる IIIF といった技術において根幹となる要素であり、利用者の利便性向上はもとより、システム間の連携や相互運用性の向上にも寄与する技術であるといえる。

（イ）パーマリンクの動向

図書館分野では、OPAC 上の書誌レコードに対してパーマリンクを付与している事例がある。国立国会

---

40 http://www.ndl.go.jp/jp/aboutus/standards/meta/about_dcndl.html （2017 年 1 月 24 日アクセス）

41 「友達の友達」という言葉が示す通り、人に関する情報を人間とコンピュータが理解できる形式で記述しようとするプロジェクトである。
http://www.foaf-project.org/ （2017 年 1 月 24 日アクセス）

42 RDF データのためのデータ・モデリング語彙。
http://www.w3.org/TR/2014/REC-rdf-schema-20140225/ （2017 年 1 月 24 日アクセス）

43 Web 上の物事の関係を一定のルールとして記述することで、コンピュータによる推論や相互利用を可能とする。RDF を拡張した語彙を使用する。
http://www.w3.org/TR/owl-features/ （2017 年 1 月 24 日アクセス）

44 http://www.ndl.go.jp/jp/aboutus/standards/meta/2011/12/ndl-term.pdf （2017 年 1 月 24 日アクセス）

45 https://dp.la/info/wp-content/uploads/2015/03/MAPv4.pdf （2017 年 1 月 24 日アクセス）

図書館の NDL-OPAC では書誌情報にパーマリンクが表示されており、このパーマリンクによる URI を用いることで書誌情報へアクセス可能である(図 24)。米国議会図書館(Library of Congress、LC)においても OPAC 上の書誌レコードに対してパーマリンクが提供されている[46]。

図 24 パーマリンクの例(NDL-OPAC の事例、赤枠部分がパーマリンク)

(ウ) パーマリンクを実現するためには

パーマリンクを実現するためには、Web 上のリソース(デジタルアーカイブ・システムにおける目録情報、デジタルコンテンツ情報等)を一意に識別するための識別子の設計が必要となる。デジタルアーカイブ・システムでは過去に作成された目録データに付与された管理番号や簿冊番号等の識別子を利用することも可能であるが、日本語文字を使う場合の制限や利便性を考えると英数字による新たな識別子の付与も検討する必要がある。

パーマリンクの技術としては、Cool URI、DOI(Digital Object Identifier)が挙げられる。

Cool URI とは、W3C が提唱するセマンティック Web を実現するための URI の付与方法であり、ガイドラインとして公開されている[47]。Cool URI では、Web 上のリソース以外の実世界のオブジェクトを識別することも考慮されており、「Web 上にあること」「曖昧でないこと」の 2 つの要件が規定されている。

DOI はコンテンツの電子データに付与される国際的な識別子であり、国際 DOI 財団がシステム及び仕様を管理し、世界で 10 機関の DOI 登録機関が DOI の登録及び管理を分担している。日本では、4 つの国内学術機関で共同運営するジャパンリンクセンター(JaLC[48])が登録及び管理を行っている。

なお、デジタルアーカイブ・システムの更新の際には、URI の維持を明示的に仕様に入れることが望まれる。やむを得ず URI を変更せざるを得ない場合でも、Web サーバの設定によりリダイレクト(ある URI へのリクエストを受け付けた際に異なる URI へ自動的に転送する)を行う等の手段によって、アクセス継続性を維持することは可能である。

---

46 http://current.ndl.go.jp/e758 (2017 年 1 月 24 日アクセス)
47 http://www.w3.org/TR/2008/NOTE-cooluris-20080331/ (2017 年 1 月 24 日アクセス)
48 https://japanlinkcenter.org/top/ (2017 年 1 月 24 日アクセス)

(4) その他の目録関連技術の動向

(ア) RiC (Records in Contexts)

RiC とは、国際公文書館会議(ICA)の Expert Group on Archival Description(EGAD)において検討が進められている、アーカイブズにおける新しい記述標準である。概念モデルである RiC-CM(Records in Context-Conceptual Model)は 2016 年 9 月にドラフトが公開され[49]、同月に開催された第 18 回 ICA ソウル大会で概要が発表された。2017 年 1 月 31 日を期限としてドラフトに対するコメントが募集されている[50]。

RiC は、概念モデルである RiC-CM と、オントロジーである RiC-O(Records in Context-Ontology)の 2 つで構成されている。RiC の概念モデルは、記述される「実体(entities)」とその「関係性(relation)」をグラフにより表現する(図 25)。ISAD(G)の「記述単位(Unit of Description)」と「多階層記述(Multilevel Description)」は、レコードは同一の方法で記述されていること、1 つの資料群(Fonds)に関する記述から始まり、階層的に記述が構成されることを前提としているが、これらの前提にあてはまらない資料が増えてきたため、RiC の概念モデルは、「レコードとレコードセット(Record and Record Set)」と「多次元記述(Multidimensional Description)」という考え方を基本としている。

RiC は、ISAD(G)を含む以下の 4 つのアーカイブズに関連する記述標準の統合を目指している。

- General International Standard Archival Description (ISAD(G))
- International Standard Archival Authority Records – Corporate Bodies, Persons, and Families (ISAAR(CPF))
- International Standard Description of Functions (ISDF)
- International Standard Description of Institutions with Archival Holdings (ISDIAH)

---

49 http://www.ica.org/sites/default/files/RiC-CM-0.1.pdf (2017 年 1 月 24 日アクセス)

50 http://www.ica.org/en/deadline-comments-ric-cm-extended-january-31-2017 (2017 年 1 月 24 日アクセス)

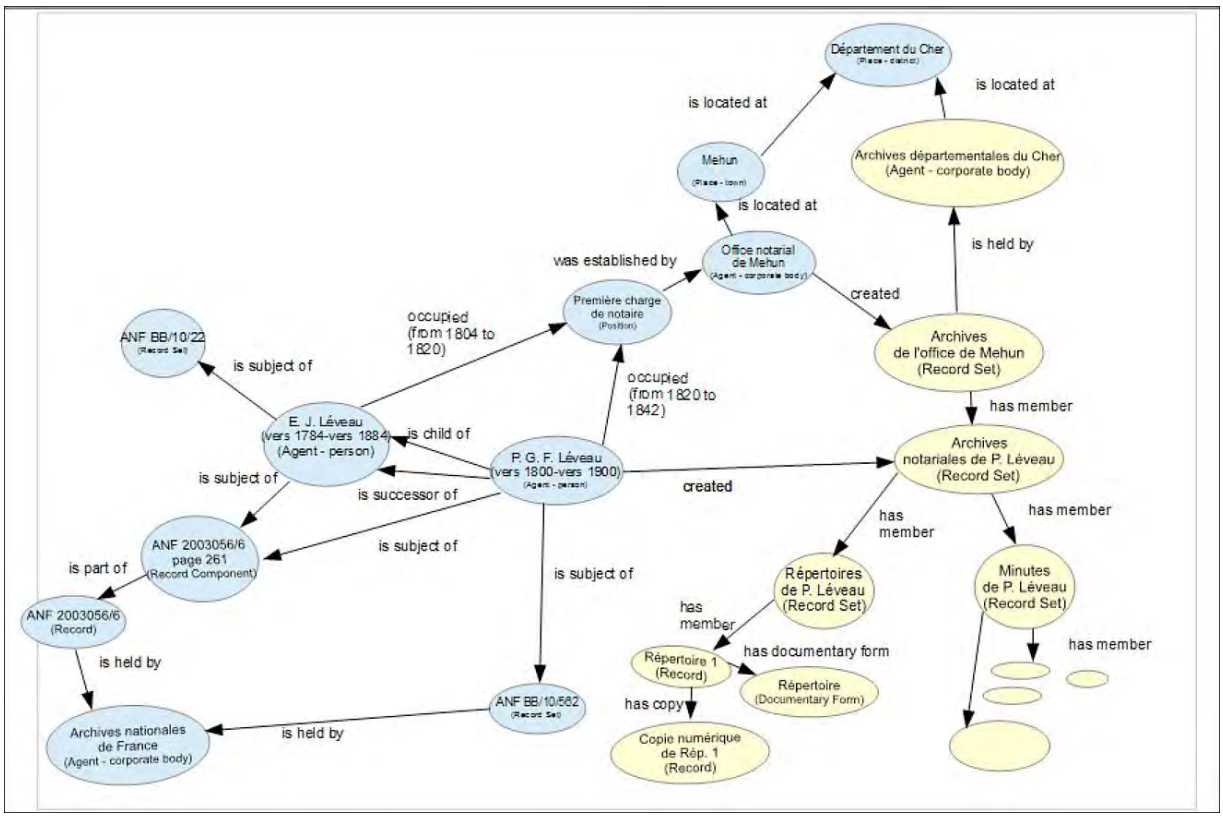


図 25 RiC-CM の概念モデル例[51]

RiC の概念モデルは、オーストラリア国立公文書館の採用しているシリーズ・システム[52]の考え方と共通している。オーストラリア・シリーズ・システムは、頻繁な組織改編とアーカイブズの記述原則の両立という問題意識を出発点とし、記述単位を独立させ、それらの関係性をリンクとして表現している。

調査時点では概念モデルのドラフトが公表された段階であるが、正式に採択されれば今後のアーカイブズの記述標準として国際的にも普及していく可能性があり、動向を注視すべきである。

（イ）メタデータに関する参考資料

本項で取り上げた目録関連技術、特にメタデータに関する考え方や用語の理解を深めるためには、一般社団法人 メタデータ基盤協議会が公開している「メタデータ情報共有のためのガイドライン[53]」が参考となる。「メタデータ情報共有のためのガイドライン」では、メタデータの提供者、利用者双方を対象とし、メタデータの設計、作成から利用、運用管理まで、メタデータの相互運用性、長期利用可能性を高めるための指針を示しており、デジタルアーカイブ・システムに関わる機関の担当者にとっても非常に有益な情報源である。

（5） まとめ

EAD に対応することで、各機関の目録情報の統一性を図ることが可能となる。事実上の国際標準であ

---

51 http://www.ica.org/sites/default/files/RiC-CM-0.1.pdf （2017 年 1 月 24 日アクセス）, P93

52 正式には“Commonwealth Record Series (CRS) System”と呼ばれている。
http://www.naa.gov.au/collection/fact-sheets/fs06.aspx （2017 年 1 月 24 日アクセス）

53 http://www.mi3.or.jp/item/A03.pdf （2017 年 1 月 24 日アクセス）

るため、国内外を問わず目録情報の交換の容易さや機械可読性が高まり、目録情報の利活用を拡大させる効果も期待される。目録情報の相互運用性を高め、二次利用を容易にするため、EAD に準拠することが推奨される。EAD はこれまでバージョンを改訂しており、現時点での最新版は EAD3 となっている。本調査の時点では EAD3 の実装例を確認することができなかったが、EAD2002 との親和性も高く、今後、記録史料を扱う機関のシステムへの実装が進むものと見込まれる。

DCMES は簡潔であることから多様な領域に適用でき、システム間連携の促進が期待できるものである。しかし一方で表現可能な情報の粒度が粗いことから、Dublin Core の拡張項目（DC タームス等の拡張された語彙）の活用による検索の精度向上も検討することが可能であると考えられる。また、横断検索以外にも、各機関がデジタルアーカイブ・システムで扱う目録データの項目を設計する際にアプリケーションプロファイルを適用することで、項目定義にかかるコストの軽減と、他機関の目録データとの共通化を図ることも可能である。

パーマリンクを用いることにより、資料目録やデジタルコンテンツの長期安定利用のために、情報へのアクセスを永続的に保証することが可能である。パーマリンクはデジタルアーカイブ・システムにおいて非常に重要な機能であると考えられる。ただし、パーマリンクを実現するには目録やデジタルコンテンツの識別子、URI を永続的に使用することを前提に設計する必要があり、その一貫性や永続性を保証するために、各機関でシステム更新を行う際に URI が維持されることを明示的に仕様として入れる等の対応が必要である。

2.2.7. 目録データの提供

目録データを二次利用可能な形で提供することにより、目録データの分析や加工等、利用者の利便性は向上する。目録データを提供する際のデータ形式、また二次利用を促進するためのオープンデータの動向について調査する。

(1) 目録データの提供

(ア) 利用者への目録データ提供

デジタルアーカイブ・システムでは、通常は Web ブラウザを通して目録データを利用者へ提供する機能を有するが、更に、表計算ソフト等によって利用者による目録データの二次利用を可能とするために、目録データを一括ダウンロードする機能があることが望ましい。

目録データを一括ダウンロードにより提供する際の形式は、CSV 等のデータ項目の意味と内容が明確に理解でき、操作性の高い形式が望ましい。

(イ) 機械可読形式でのデータ提供

一方で、他システムによるサービスやアプリケーションが目録データを利用可能とすることにより、目録データの利活用の機会を拡大することが可能となり、目録データの相互運用性を向上させるものとなる。他システムのサービス等が目録データを利用可能とするに当たっては、目録のデータ項目の意味や値を他システムが理解できる形式で提供することが求められる。

Web 上で普及している機械可読なデータ形式として RDF(Resource Description Framework)がある。RDF は、Web 上のあらゆる情報を機械可読にする記述方式で、すべてのデータが Subject(主語)、Predicate(述語)、Object(目的語)という 3 つの組(トリプル)で構成されたグラフとして表現される。

RDF により記述されたデータを蓄積する RDF ストアからデータを検索するためのクエリー言語として SPARQL(SPARQL Protocol and RDF Query Language)が World Wide Web Consortium(W3C)により標準化されている[54]。RDF 化された目録データを他システムにより利用可能とするためにはこれらの仕組みを用いることとなる。

(2) オープンデータ

公共データの機械判読可能な形式での提供については、国によるオープンデータの推進においても求められている。平成 24 年に政府の IT 総合戦略本部が「電子行政オープンデータ戦略」を発表[55]しており、デジタルアーカイブ・システムで提供する目録データやデジタルコンテンツも、今後より幅広く利用者へ提供していくことが必要になると考えられる。

なお、オープンデータについては Open Knowledge International が公開している"Open Data Handbook"[56]が参考となる。"Open Data Handbook"ではオープンデータとは何か、なぜオープンデータが必要なのか、オープンデータとするために何をしなければならないのかといったことが解説されており、デジタルアーカイブ・システムに関わる機関の担当者にとっても非常に有益な情報源である。

---

54 http://www.w3.org/TR/sparql11-overview/ (2017 年 1 月 24 日アクセス)

55 http://www.kantei.go.jp/jp/singi/it2/denshigyousei.html (2017 年 1 月 24 日アクセス)

56 日本語版が次の URL で提供されている。http://opendatahandbook.org/guide/ja/ (2017 年 1 月 24 日アクセス)

（ア） 我が国におけるオープンデータの取り組み

オープンデータの取り組みの一つである「二次利用を促進する利用ルールの整備」に関しては、「世界最先端IT国家創造宣言工程表」及び「電子行政オープンデータ推進のためのロードマップ」（ともに平成25 年 6 月IT総合戦略本部決定）、「二次利用の促進のための府省のデータ公開に関する基本的考え方（ガイドライン）」（平成 25 年 6 月各府省 CIO 連絡会議決定）で基本的考え方が示されている。この考え方に基づき、各府省 Web サイトの利用規約のひな形として策定されたのが「政府標準利用規約(第 1.0 版)」である。平成 27 年 11 月 24 日の各府省 CIO 連絡会議において決定された第 2.0 版では、国際的にオープンなライセンスとして通用するように「クリエイティブ・コモンズ・ライセンスの表示 4.0 国際(CC BY)」との互換性が明記された。[57]

国内のデジタルアーカイブ・システムでは、京都府立京都学・歴彩館(旧京都府立総合資料館）が東寺百合文書WEB上のすべてのコンテンツを原則として「クリエイティブ・コモンズ 表示 2.1 日本 ライセンス(CC BY 2.1 JP)」で提供している例[58]があるが、二次利用に関しては各機関独自の方針に従っているのが現状である。

（イ） 国外におけるオープンデータへの取り組み

オープンデータへの取り組みは主に欧米を中心として海外において積極的に進められている。

米国では、政府の信頼性向上、民主主義の強化、政府の効率性と有効性の向上を掲げ、データポータルサイトである「Data.gov」を開設している。

英国でも 2010 年に「data.gov.uk」を開設し、再利用可能かつ機械判読可能な形でデータの公開を行うこと、同一のオープンライセスで公開し、営利目的も含め自由に活用可能にすること、使いやすい単一のオンラインアクセスポイントでデータ入手可能にすること等を掲げた「透明性原則」を発表している。

（ウ） 技術要素

オープンデータでは、「5 つ星スキーム(5 Star Open Data)」と呼ばれる公開度に応じた評価指標が Open Data Institute により提唱されている[59]。4 つ星のオープンデータでは URI を識別子として使用すること、5 つ星のオープンデータではさらに他のデータへのリンクを含むことが必要となる。4 つ星のオープンデータを実現する技術として RDF、5 つ星のオープンデータを実現する技術として Linked Open Data(LOD)がある。

---

57 http://www.kantei.go.jp/jp/singi/it2/densi/kettei/gl2_betten_1_gaiyou.pdf （2017 年 1 月 24 日アクセス）

58 http://hyakugo.kyoto.jp/guide （2017 年 1 月 24 日アクセス）

59 http://5stardata.info/ja/ （http://5stardata.info/en/ の日本語訳） （2017 年 1 月 24 日アクセス）

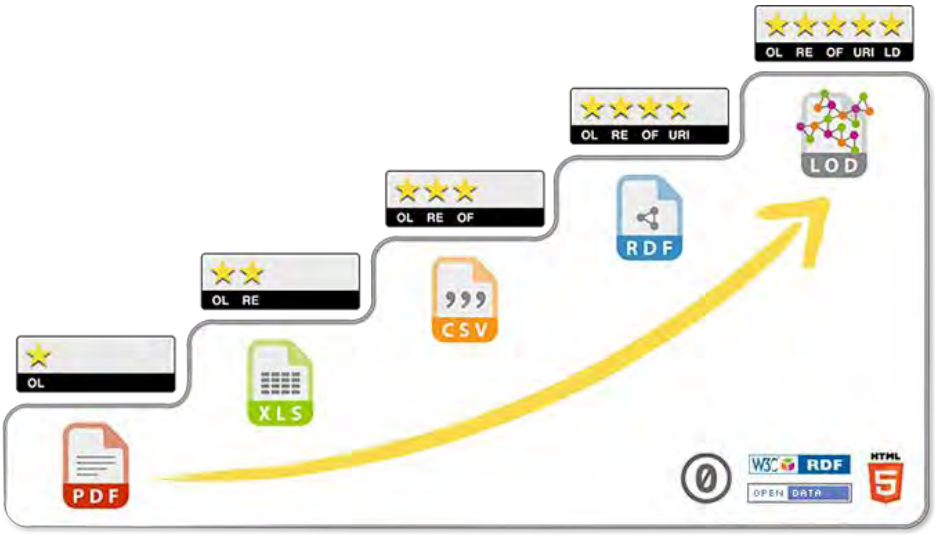


図 26「5 つ星スキーム」のイメージ

RDF はデータを「グラフ」として表現する仕組みであり、機械可読性は高いがデータ設計には高度な専門知識が必要となる。さらに LOD ではどのように、どのデータへリンクさせるかを考慮しなければならない。一般社団法人メタデータ基盤協議会が運営する「Meta Bridge」[60]では、メタデータの相互運用性を高めるために様々な分野のメタデータスキーマを登録して公開することができるため、RDF や LOD のデータ設計において有用である。

（エ）ライセンス

オープンデータでは、技術的にオープンにするだけでなく、法的にオープンとすることが必要となる。"Open Data Handbook"では、データに対する知的財産権が存在し、第三者によるデータの利用や再利用、再配布などに明示的な許可を要する場合でも、知的財産権が存在するかどうかが不確かな場合でも、何らかのライセンスを適用して明確にしておくことが求められている[61]。

オープンデータに適用するライセンスとして、Open Knowledge International はオープンデータに準拠した推奨ライセンスのリストを公開している[62]。また、オープンデータにライセンスを適用する際の簡単な手引きが"Open Data Commons"のサイトで公開されている[63]。

（オ）識別子の標準化

オープンデータ化を推進するにあたり、識別子の国際標準化も進められている。代表的な識別子に「図書館及び関連組織のための国際標準識別子(International Standard Identifier for Libraries and Related Organizations; ISIL)」がある。ISIL は、図書館および関連組織(公文書館や博物館など)に割り当てられる国際的な識別子(ID)であり、ISO15511 として規定されている[64]。一方で、公文書館等のアーカイ

---

60 https://www.metabridge.jp （2017 年 1 月 24 日アクセス）
61 http://opendatahandbook.org/guide/ja/how-to-open-up-data/ （2017 年 1 月 24 日アクセス）
62 http://opendefinition.org/licenses/ （2017 年 1 月 24 日アクセス）
63 http://opendatacommons.org/guide/ （2017 年 1 月 24 日アクセス）
64 http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=57332

ブズ所蔵機関に関わる識別子については「アーカイブズ所蔵機関の記述に関する国際標準（ISDIAH）」が国際公文書館会議(ICA)により公開されている[65]。「アーカイブズ所蔵機関の記述に関する国際標準（ISDIAH）」では、「各国においては、アーカイブズ所蔵機関それぞれに対して固有の識別子を定め、維持することが望まれる。また、国際レベルで開発された文化機関を体系化する他のシステムと一貫していることが望ましい。」とされており、ISIL も含め言語や国名といった各種国際標準が関連標準として示されている。

(3)　まとめ

デジタルアーカイブ・システムでは、目録データの提供方法として、利用者が Web ブラウザ上で目録を閲覧可能とすることに加え、CSV 等の形式による一括ダウンロードを可能とすることが望ましい。

また、デジタルアーカイブ・システムで提供するデータは、今後より一層オープンデータへの対応が求められると考える。オープンデータ化に当たり、機械可読性の高い RDF 等のデータ形式による目録データの提供や LD(Linked Data)もしくは LOD(Linked Open Data)への対応が考えられる。

オープンデータ化の推進において、データの提供形態の周知による、データ利用者の利便性向上の観点からデータの利用規約の明示が望ましいと考えられる。データの利用規約に当たっては、政府標準利用規約やクリエイティブ・コモンズといった標準を踏まえつつ、各機関の事情を考慮して検討する必要があると考えられる。

（2017 年 1 月 24 日アクセス）

[65] http://www.ica.org/sites/default/files/CBPS_2013_guidelines_ISDIAH_JPN.pdf （2017 年 1 月 24 日アクセス）

2.2.8. 情報検索技術

目録データを高速に検索するための検索エンジン技術について、また複数の公文書館等の資料を一元的に検索する横断検索技術について調査する。

(1) 検索エンジン技術

情報検索システムとは、検索対象のデータ及びそのデータに対するメタデータを蓄積するデータベースと、利用者からのデータベースへの問い合わせ(検索リクエスト)を受け付け、問い合わせ先に検索結果を送り返す(レスポンス)ための仕組みにより構成されるシステムである。

デジタルアーカイブ・システムにおける検索は資料の目録データに対する操作が基本となる。検索のための処理(検索アルゴリズム)を担う仕組み(検索エンジン)は、利用者が求める資料へ容易に辿り着くための重要な機能である。「検索エンジン」という言葉は、ロボット型検索エンジン、ディレクトリ型検索エンジンといったインターネット上の情報を検索する機能・サービス、ソフトウェアを指す場合もあるが、本項では、デジタルアーカイブ・システムで主に目録情報(構造化データや XML、JSON[66]のような規則性を持った非構造化データであらわされるメタデータ)を対象として検索を行うソフトウェアである「全文検索エンジン」を対象とし、技術的な最新動向を調査した結果を扱う。また、新しいデータベース技術である「NoSQL」の動向についても触れる。

(ア) 代表的な全文検索エンジンソフトウェアと動向

全文検索エンジンとして代表的なソフトウェアとして、「Lucene」[67]、「Solr」[68]、「Elasticsearch」[69]がある。いずれもオープンソースであり、ライセンスは Apache License 2.0 の元で公開されている。Lucene は 1000 万ドキュメント程度のデータを高速に検索できる性能を持っており、ワイルドカード、正規表現、あいまい検索、範囲検索、フィルタリング、ファセット表示といった多彩な検索機能を利用できるが、これらの特徴は Solr、Elasticsearch にも引き継がれている。

Elasticsearch は検索エンジンのみであれば無償で使用可能である。(Elastic 社による有償サポートあり)。また、クラウド上で Elasticsearch を利用できるサービスが提供されている[70]。

---

66 JavaScript Object Notation の略。JavaScript 言語のオブジェクト表記法をベースとしたデータ記述言語であり、テキストフォーマットである。

67 https://lucene.apache.org/ (2017 年 1 月 24 日アクセス)

68 http://lucene.apache.org/solr/ (2017 年 1 月 24 日アクセス)

69 https://www.elastic.co/jp/products/elasticsearch (2017 年 1 月 24 日アクセス)

70 例えば Amazon Elasticsearch Service が挙げられる。
https://aws.amazon.com/jp/elasticsearch-service/ (2017 年 1 月 24 日アクセス)

表 13 代表的な全文検索エンジンソフトウェア

検索エンジン	特徴	ライセンス
Lucene	Apache プロジェクトで開発されている全文検索エンジンライブラリ。	オープンソース(Apache License 2.0)
Solr	Lucene に管理画面やキャッシュ機構といったアプリケーションとしての機能を付加した Apache プロジェクトの全文検索エンジン。	オープンソース(Apache License 2.0)
Elasticsearch	Elastic 社が Lucene をベースとして分析機能や分散機構を付加した全文検索エンジン。	オープンソース(Apache License 2.0)

(イ) RDBMS での全文検索

データベースシステムでは、リレーショナルデータベース(RDB)と呼ばれる表とその関係を表したモデルによるデータベースと、その管理システムである RDBMS(リレーショナルデータベース管理システム)が広く用いられている。

RDBMS は基本的な機能として全文検索に対応するものではないが、商用 RDBMS である Oracle 等、全文検索に対応する機能を有するものが出てきている。

(ウ) データベース技術の動向

近年では RDB のような固定されたスキーマを必要とせず、膨大な非構造化データである「ビッグデータ」を扱うことに特化した「NoSQL(Not only SQL)」と呼ばれるデータベースも利用されている。NoSQL は RDBMS 以外のデータベース管理システムを指す言葉であり、キー・バリュー型、カラム指向型、グラフ型、ドキュメント指向型といった分類がある。

NoSQL は SNS でアクセスログやテキスト、動画、画像といった多様なデータをリアルタイムに処理し、分析することで企業のマーケティングに活用されているケースがある。しかし、リアルタイムでの処理速度を優先し RDB の持つ排他制御の仕組みを簡素化しているため、データの整合性やトランザクションが必要な業務処理には向いていない。

現状では NoSQL は目録という構造化データを管理する業務を含むデジタルアーカイブ・システムには適していないと考えられるが、今後技術発展により RDB との相互補完を果たす仕組みとして実用化が進むことが期待されるため、動向を注視すべきである。

(2) 横断検索技術

横断検索は、横断検索のポータルを提供する機関(国立公文書館等)から、他の機関の目録データベースを一定の規約(検索プロトコル)に則って同時かつ横断的に検索するものである。また、一度の検索で、連携している複数の機関のデジタルアーカイブ・システムを横断的に検索できるため、利用者の利便性向上に大きく資する機能である。

横断検索による連携には、検索する側の機関には検索プロトコルに対応したクライアント機能が必要であり、検索される側の機関には同じ検索プロトコルに対応したサーバ機能や API が必要となる。

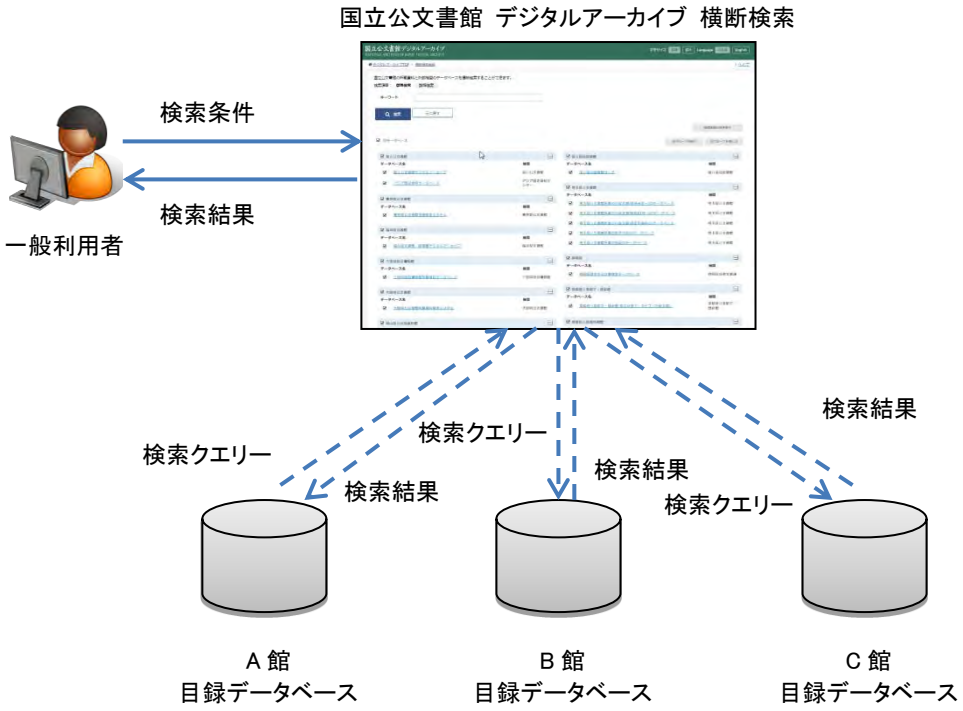


図 27 デジタルアーカイブ・システムにおける横断検索イメージ(国立公文書館の例)

本項では、横断検索技術の中核である検索プロトコルの代表的なものについて調査する。

(ア) 代表的な検索プロトコル

横断検索のための代表的な検索プロトコルとしては、Z39.50、SRU/W、OpenSearch を挙げることができる。表 14 に各プロトコルの概要及び特徴を示す。

表 14 代表的な検索プロトコル

プロトコル	概要、特徴
Z39.50	Z39.50 は、データベースに対するクライアント・サーバ型の検索プロトコルである。ANSI/NISO 標準(ANSI/NISO Z39.50-2003)および ISO 国際標準(ISO 23950)となっており、米国議会図書館が標準の管理を行っている[71]。主に図書館の目録システムにおける横断検索用途で利用されている。 Z39.50 はウェブが普及する前に策定されたプロトコルであり、現在はウェブベース(HTTP プロトコル)の SRU/W に置き換わりつつある。 現在の最新版は 2003 年版であり、仕様の改訂は行われていない。
SRU/W SRW(Search/Retrieve Web Service) SRU(Search/Retrieve via URL)	SRW(Search/Retrieve Web Service)は、ウェブベースの検索プロトコルである。SOAP 通信により検索条件を XML 形式で記述して送信し、検索結果を XML で取得する。 SRU(Search/Retrieve via URL)は、REST により URL 内に検索条件を埋め込んでアクセスし、検索結果を XML で取得する。 SRW と SRU は通信方式の差異はあるが、検索プロトコルとしては基本的に同一である。検索クエリーには CQL(Contextual Query Language)と呼ばれるクエリー言語を用いる。[72] SRU、SRW のいずれも米国議会図書館が標準の管理を行っている[73]。 現在の最新版は SRU2.0[74]であり、OASIS の"searchRetrieve 1.0"標準の一部となっている。
OpenSearch	OpenSearch は Amazon.com の子会社 A9 によって開発された検索エンジン提供者のための技術である。OpenSearch 仕様はクリエイティブ・コモンズ・ライセンス下で、A9 によりライセンスされている[75]。 OpenSearch はウェブベースでの横断検索が実現可能な仕組みであり、検索エンジンを特定し、説明するための OpenSearch Description ファイル、検索結果を供給するための形式である OpenSearch RSS/Response、検索結果を表示するサイトである OpenSearch Aggregators により構成される。多くのソフトウェアやライブラリでサポートされている[76]。 現在の最新版は OpenSearch 1.1 (Draft 5)である。

Z39.50 は図書館目録に特化した仕様となっており、クライアント・サーバ型であることから用途は制限される。SRU/W、OpenSearch はいずれも HTTP をベースとしており(ウェブベース)、汎用性が高く開発が

---

71 http://www.loc.gov/z3950/agency/ (2017 年 1 月 24 日アクセス)
72 http://www.loc.gov/standards/sru/cql/index.html (2017 年 1 月 24 日アクセス)
73 http://www.loc.gov/standards/sru/ (2017 年 1 月 24 日アクセス)
74 http://www.loc.gov/standards/sru/sru-2-0.html (2017 年 1 月 24 日アクセス)
75 http://www.opensearch.org/Home (2017 年 1 月 24 日アクセス)
76 http://www.opensearch.org/Community/OpenSearch_software (2017 年 1 月 24 日アクセス)

容易である。

（イ）Web スクレイピング

Web 上での横断検索を実現する方法として、Web スクレイピングがある。Web スクレイピングは、検索相手側の Web ページの HTML データを収集してデータを抽出し、整形する技術である。検索の相手側では横断検索のためのサーバ機能や API への対応が不要であるが、検索側は相手側の HTML データに合わせたデータ抽出が必要となるため、サイトの更新や仕様変更への対応が難しいという側面がある。

Web スクレイピングの特徴的な事例として、株式会社カーリルが運営する図書館の蔵書検索サービス「カーリル」がある。カーリルでは、全国 6,000 以上の公共図書館、大学図書館、専門図書館を対象に、蔵書の横断検索と蔵書の貸出状況をリアルタイムに取得できる。多くの図書館の蔵書検索システム(OPAC)を検索し、その結果として返される HTML データを Web スクレイピングにより解析することで横断検索を実現している。

(3)　その他データ提供技術

コンピュータシステム間で、ある処理を呼び出すため呼び出し方法や受け渡すデータの形式等のインターフェース仕様を定めたものをアプリケーション・プログラミング・インターフェース(Application Programming Interface、API)という。

現在、インターネット上でサービスを提供するウェブベースのシステムが広く普及し、HTTP 上で API を実現する Web API が様々なサービスで利用されている。Web API は HTTP の GET や POST といったリクエストを使い、XML や JSON といった標準的なフォーマットでデータを返す仕組みであることが多い。よく利用されている Web API の例として、Google が提供している Google Maps API がある。Google Maps API を利用することで、自機関のデータベースのデータを Google が提供する地図上にプロットして表示するといったことが容易に実現できるようになる。

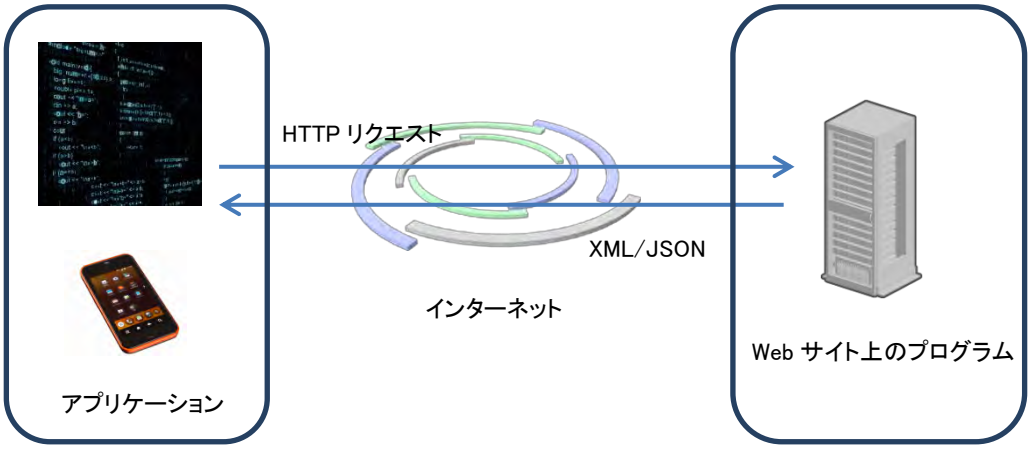


図 28 Web API のイメージ

「(2)横断検索技術」で調査した SRU や OpenSearch、「2.2.7 目録データの提供」で調査した SPARQL も、HTTP で検索のリクエストを受け、検索結果を XML 形式のデータで返すという点では Web API の一種である。

デジタルアーカイブ・システムにおいても、オープンデータ推進の観点では、機械可読形式によるデータ提供と合わせて、Web API を提供することにより、利用者や他機関が自分のアプリケーションから各機関の目録データやデジタルコンテンツデータを利用することが可能となり、利活用の拡大が期待できる。

(4) まとめ

デジタルアーカイブ・システムでは目録情報に非構造化データを扱うことも想定される。そのための対応として、目録検索における全文検索に対応した技術(全文検索エンジン)を用いることが検討可能である。全文検索エンジンにはオープンソースのソフトウェアのほか、RDBMS の全文検索機能を利用する方法も考えられる。

また、横断検索に当たっては、HTTP ベース(ウェブベース)の検索プロトコルである SRU/W、OpenSearch のいずれかを実装することが、相互運用性の向上という点からも推奨される。Web スクレイピングはこれらのプロトコルをサポートしていない機関でも横断検索対象とできるが、横断検索を行う側の機関で各機関の HTML を解析し、横断検索に係る処理を実施するための仕組みが必要となり、横断検索を行う側の機関における負担が大きく、相互運用性が低いものである。その他のデータ提供技術としては、オープンデータのために WebAPI を提供することが推奨される。

2.2.9. デジタルコンテンツ配信技術

特殊なソフトウェアを必要とせずに、デジタルアーカイブ・システムの検索結果から連動してデジタルコンテンツを閲覧できること、またデジタルアーカイブで扱うコンテンツの多様化を踏まえて、画像に限らず、動画や音声資料等マルチメディアの配信に対応することは、利用者の利便性向上につながる。デジタルコンテンツのフォーマット、配信方法、および配信にかかる有益な技術について調査する。

(1) デジタルコンテンツフォーマット

デジタルアーカイブ・システムにおけるデジタルコンテンツの配信方法の検討のため、画像、動画、音声等のマルチメディアコンテンツのファイルフォーマットについて調査する。また、利用者向けのデジタルコンテンツの配信フォーマットが保存フォーマットと異なる場合には、フォーマット変換の必要があることから、一部保存フォーマットについても触れる。

(ア) 画像フォーマット

配信用の画像フォーマットは JPEG が主流である。また、JPEG による画像データを埋め込んだ PDF はマルチページに対応可能であり、その可用性は国際規格として担保されている。保存用の画像フォーマットとしては TIFF、BMP、JPEG2000 が使われている。JPEG2000 は国際標準であり、高圧縮率、高画質で有用性は高いが、デジタル化のコストや汎用性、デジタル化の量や保存期間等を考慮する必要がある。

(イ) 動画フォーマット

動画フォーマットには、動画と音声の符号化方式(コーデック)とそれらをまとめるコンテナという 2 つの要素がある。ここでいう「動画フォーマット」はコンテナを指す。コーデックには H.264、MPEG-4、MPEG-2、DivX 等様々な種類がある。

現在主流の動画フォーマットは AVI、MP4 である。AVI は、Microsoft が開発した古い動画フォーマットであるが、コーデックを選ばないため、現在も利用され続けている。

MP4 は MPEG-4 Part14 の略称で、Apple のメディア技術 QuickTime のファイル形式を元に、ISO/IEC の動画関連技術の国際標準規格である MPEG-4 の第 14 部で策定された動画フォーマットである。MP4 のコーデックには主に H.264 が使用されており、速いスピードと高い品質のため、インターネットにおけるメディア配信に最適な形式として普及している。MP4 は、スマートフォンでは iPhone 系と Android 系の全てでサポートされている。また、「HTML5 の video タグの MP4 対応」については、主要ブラウザがサポートしているため、配信する上では MP4 が適切と考えられる。

動画の保存と配信は同じフォーマットでよいが、動画の高精細化に伴い新しい圧縮技術も開発され続けているため、今後の動向について注視する必要がある。

(ウ) 音声フォーマット

音声フォーマットも動画と同様に、音声の符号化方式(コーデック)とそれらをまとめるコンテナという2つの要素があるが、1 つのコーデックが 1 つの音声フォーマットに対応することが多い。

現在主流の音声フォーマットは WAV、AAC、MP3 である。

WAV は、Microsoft と IBM により開発された音声データ記述のためのフォーマットであるが、通常は無圧縮のためファイルサイズが大きくなる。

AAC は、1997 年に出来た規格である。地上デジタル放送や BS デジタル放送の音声は、この形式が使われている。iPod や iPhone でよく使用されるフォーマットであるが、Android は標準では対応していない。

MP3 は、20 年以上前にできた古くからある規格であり、主要な OS や主要なブラウザが対応している非可逆圧縮[77]音源の標準フォーマットとなっている。よって、配信する上では MP3 が適切と考えられる。音声の保存フォーマットは、配信と同じフォーマットでよい。

(2) デジタルコンテンツの配信

デジタルコンテンツの配信方法について、利用者の環境においてストレスなくコンテンツを閲覧するための技術とその動向を調査した。

(ア) ダウンロード方式

ダウンロード方式は、Web サーバから対象となるファイルを HTTP または HTTPS 上で転送する方式である。フォーマットを問わず配信できるが、ネットワークを効率的に利用できないため、データサイズが大きいファイルは転送時間が長くなる。

(イ) ストリーミング方式

ストリーミング方式は、データサイズが大きい動画・音声に適した配信方式である。ストリーミング方式では必要なデータを都度転送するため、ネットワークを効率的に利用でき、利用者も指定した位置からの再生が容易に行える。但し、配信側にはストリーミングに対応した配信用ソフトウェアを導入する必要があり、利用者側にはストリーミングに対応した再生ソフトウェア等が必要な場合がある。

ダウンロード方式とストリーミング方式の中間的な方式として、プログレッシブダウンロードがある。プログレッシブダウンロードは、サーバから利用者の環境に動画等のデータをダウンロードしながら再生を行う方式である。ファイル全体のデータをダウンロードすることはダウンロード方式と変わらないため、ネットワークを効率的に利用できない点はダウンロード方式と同じだが、Web サーバの設定のみで対応でき、利用者は全データのダウンロードが完了しなくても途中から再生を行える。

コスト面では、プログレッシブダウンロードの場合は Web サーバの設定作業のみで実現できるが、ストリーミングを実現するには、有償のソフトウェア導入やサーバ増強が必要な場合があるため高くなる。

(3) デジタルコンテンツの配信に関わる有益な技術

本項では、利用者の利便性向上に資する技術として、国際的な画像配信フレームワーク IIIF、コンテンツキャッシュ技術について調査する。

(ア) IIIF[78](International Image Interoperability Framework)

IIIF は、画像へのアクセスを標準化し、相互運用性を確保することを目的とした国際的な画像配信のた

---

77 多少のデータの欠落を許容する代わりに、劇的に圧縮効率を高めた圧縮方式。非可逆圧縮方式により圧縮されたデータを復号(伸長)しても、圧縮前のデータを完全に復元することはできない。

78 http://iiif.io/ (2017 年 1 月 24 日アクセス)

めのフレームワークである。英国図書館、フランス国立図書館、オックスフォード大学ボドリアン図書館、スタンフォード大学、プリンストン大学、イェール大学などを中心に策定された。Image API、Presentation API、Search API という 3 つの API の仕様が定められ、公開されている。デジタルアーカイブ・システムにおいても画像配信の技術として IIIF を採用している事例が見られる。

配信する側が IIIF に対応することで、利用者はビューアを自由に選択することが可能となる。また、デジタルアーカイブ・システムと他システム間での画像データ連携が容易となることが期待できる。

以下に IIIF で定義されている API の概要を記載する。

- IIIF Image API 2.1[79]

この API では、標準的な HTTP や HTTPS のリクエストに応答して画像を返す Web サービスの仕様を定めている。領域、サイズ、回転、品質特性およびリクエストされる画像フォーマットを URI で指定できる。画像に関する基本的な技術情報をリクエストするように URI を構成することができるため、ビューア等でそれらの情報を利用することができる。この API は、文化遺産機関が管理しているデジタル画像リポジトリの画像資源を体系的に再利用可能とするために考案されたものだが、どのような画像リポジトリやサービスにも採用できる。

- IIIF Presentation API 2.1[80]

この API では、絵画等の単一ビューをはじめ、写真の両面、彫像の基本 4 面ビュー、1 冊の新聞や図書の多数のページなど、複数の画像から構成される集合体の構造やレイアウトを、標準的な方法で提供する方法を記述している。コレクションや提供機関にまたがるコンテンツを利用して、豊かでダイナミックな経験を実現するための様々なスタイルのビューアを実装するための仕様である。この API の主な要件は、ビューの順序、ビューの表示に必要なリソース、何が表示されているのかをユーザーが理解できるようにするために必要な記述情報を提供することである。

- IIIF Search API 1.0.0[81]

この API ではでは、マニフェスト(Manifest)、範囲(Range)、コレクション(Collection)などの、1 つの IIIF 資源内の注釈コンテンツの検索を行うための仕様を定めている。注釈コンテンツとは、画像内のテキスト情報や翻訳テキスト、ユーザーが提供したコメント情報を指し、メタデータは対象外となっている。また、未知のコンテンツの検索を容易に行えるよう、検索語の自動補完(オートコンプリート)に関連するサービスが定義されている。

（イ）コンテンツキャッシュ技術

コンテンツキャッシュ技術の代表的なものとして、コンテンツデリバリネットワーク(Content Delivery Network, CDN)と呼ばれるものがある。CDN は、地理的に分散された複数のサーバ(エッジサーバ)に同一のコンテンツを配置し、利用者からのアクセスを分散させることで、コンテンツ提供者のサーバ(オリジ

---

79 http://iiif.io/api/image/2.1/ （2017 年 1 月 24 日アクセス）
80 http://iiif.io/api/presentation/2.1/ （2017 年 1 月 24 日アクセス）
81 http://iiif.io/api/search/1.0/ （2017 年 1 月 24 日アクセス）

ンサーバ)の負荷軽減と Web サイトの表示速度向上を実現する仕組みである。

CDN はその仕組み上、静的コンテンツ(HTML、画像、CSS 等)の配信に有効であるが、検索結果などの変動要素を含む動的コンテンツの配信には向かない。また、静的コンテンツでも更新頻度が高い場合は、ページの有効期限を考慮する必要がある。

主要な CDN サービスの事例を表 15 に示す。

表 15 主要な CDN サービス

サービス名	概要、特徴
Akamai[82]	米国の Akamai Technologies 社が提供する、世界最大の CDN。インターネット通信量の 15～30%を処理しているとされており、その要求を適切に処理するために、72 か国に 84,000 台以上のサーバを設置している。 料金体系は非公開となっている。
Amazon CloudFront[83]	Amazon Web Service の 1 つとして提供される CDN。 データ転送量(GB 単位)またはリクエスト数による従量課金となっている。
Cloudflare[84]	米国 Cloudflare 社が提供する CDN。 サポートされる機能や規模に応じて、無料から企業向けまで 4 つの価格体系がある[85]。

(4) まとめ

デジタルコンテンツの配信にはダウンロード方式とストリーミング方式がある。ダウンロード方式はデジタルコンテンツをファイル形式で転送する方式であり、ストリーミング方式は必要なデータを都度転送する方式である。また、ダウンロード方式とストリーミング方式の中間的な方式として、プログレッシブダウンロード方式がある。プログレッシブダウンロード方式は、サーバから利用者の環境に動画等のデータを都度転送しながらファイル形式でダウンロードすることにより、ダウンロード途中であってもデータの再生等利用が可能な方式である。

配信対象となるデジタルコンテンツの権利等管理の観点からは、ストリーミング方式による対応となるが、配信用のサーバソフトウェアの導入やサーバの機能面での増強整備が必要となり、コスト負担について懸念されるところである。

また、画像コンテンツにおいては、IIIF に対応した画像配信が普及しつつある。画像配信を IIIF に対応することで、配信する画像コンテンツの相互運用性を高めることが可能であり、デジタルアーカイブ・システムへの導入は、利用者が使用したい画像閲覧環境を選択可能となり、利便性の向上や利用者の拡大を図ることができるものと考えられる。

コンテンツキャッシュ技術ではサーバの負荷軽減や表示速度の向上が期待できる。コンテンツキャッシュ技術の導入により、更新頻度が低く、ファイルサイズが大きい静的なコンテンツ(画像、動画、音声等)の配信において、配信速度の向上が見込まれる。但し、有償のサービスが多く、1 日当たりのページビューが数千件程度であれば費用対効果は低い。想定されるアクセス数に応じて検討する必要がある。

---

82 https://www.akamai.com/ (2017 年 1 月 24 日アクセス)
83 https://aws.amazon.com/jp/cloudfront/ (2017 年 1 月 24 日アクセス)
84 https://www.cloudflare.com/ (2017 年 1 月 24 日アクセス)
85 https://www.cloudflare.com/plans/ (2017 年 1 月 24 日アクセス)

2.2.10. クライアント環境

クライアント環境は、デジタルアーカイブ・システムを利用する利用者の環境のことである。幅広い利用者にとって使いやすいデジタルアーカイブ・システムとするために、クライアント環境における技術動向を把握する。ブラウザ等ソフトウェア技術や、PC やスマートデバイスにおける Web ページ等再生環境をはじめ、ビューア、クライアント環境におけるネットワーク、Web アクセシビリティ等を調査する。

(1) ブラウザ等ソフトウェア技術

デジタルアーカイブ・システムはインターネットを通じて目録や画像を提供するものである。幅広い利用者が閲覧でき、使い易さや操作性に配慮した仕組みを、ブラウザを通して実現することが求められる。

かつては使い易さや操作性の高いユーザーインターフェースを実現するために、Flash やプラグインソフトウェアを用いることが主流であったが、近年では主要なブラウザで HTML5 や CSS3/4(Cascading Style Sheets, level 3/4)への対応が進み、ブラウザのみで高機能なユーザーインターフェースを作ることができるようになった。また、HTML5 や CSS3/4 を使用することにより、動画・音声データの取り扱いが容易になった。

プラグインソフトウェアの導入は利用者の環境に依存し、導入の手間やセキュリティの問題を含む可能性もあるため、利用促進の妨げとなる恐れがある。デジタルアーカイブ・システムでも扱われる PDF 形式のデータは、プラグインソフトウェアを使わなければ閲覧できなかったが、プラグインソフトウェアを使わずに閲覧することができるブラウザも出現してきており、幅広い利用者の閲覧に寄与している。

また、JavaScript により非同期通信を行う Ajax と呼ばれる技術が普及しており、ブラウザのみでデスクトップアプリケーション並みのユーザーインターフェースや機能を、低い開発コストで実現できるようになっている。

ただし、ブラウザのバージョンアップサイクルは早く、旧バージョンで利用可能である機能が新バージョンでは実装対象外となることもある。

(2) PC 等ハードウェアへの対応技術(スマートデバイスへの対応を含む)

利用者の環境に合わせ、提供可能なサービスや利便性への影響を検討するため、PC やスマートデバイス(スマートフォン、タブレット)ごとのシェア及び技術的特徴を調査する。

(ア) PC、スマートデバイスにおける Web ページ等再生環境

全世帯の「主要耐久消費財等の普及・保有状況」によると、平成 28 年 3 月末現在、タブレットについては 26.2%、スマートフォンについては 56.6%の普及率であり[86]、図 29 の通りスマートデバイスの出荷台数は PC を上回っている。

デジタルアーカイブ・システムでも、PC による閲覧に加えて、今後はスマートデバイスによる閲覧が増加すると推測され、マルチデバイスに対応することが求められる。

---

[86] 総務省統計局『消費動向調査』
http://www.e-stat.go.jp/SG1/estat/NewList.do?tid=000001014549 (2017 年 1 月 24 日アクセス)

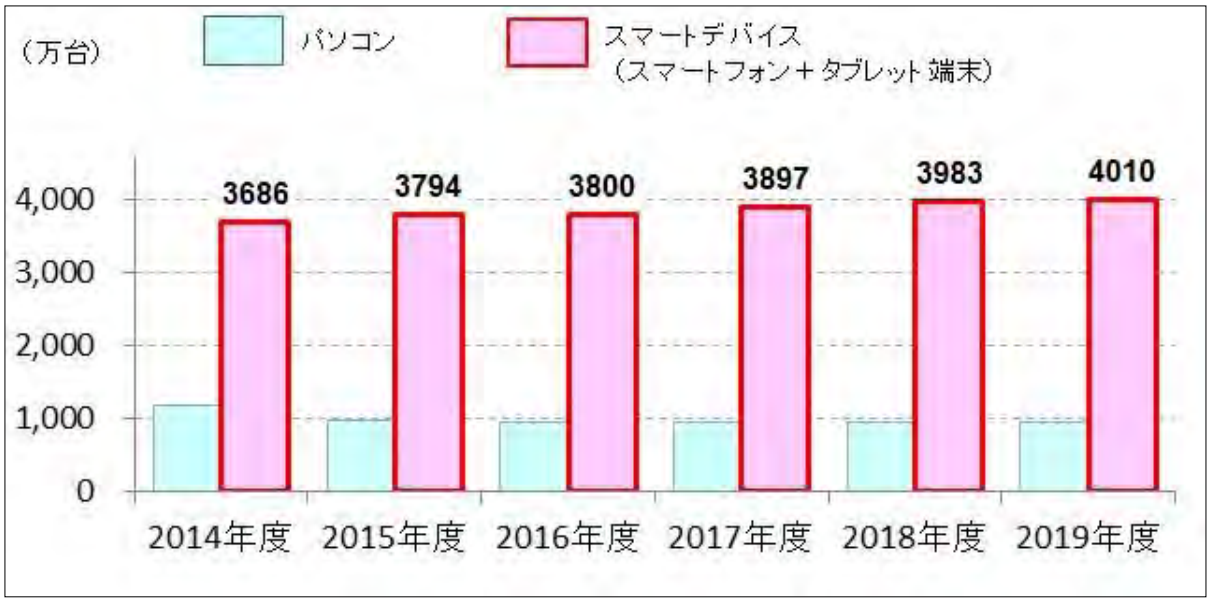


図 29 スマートデバイスとパソコンの国内出荷台数推移[87]

スマートデバイスは、従来の PC とは異なるディスプレイサイズ・解像度を持ち、またタップやスワイプといった PC には無い操作がある。利用者拡大のためには、これらに適したユーザーインターフェースが求められる。

（イ）レスポンシブウェブデザイン

多様な利用者の環境に応じたユーザーインターフェース構築手法の一つとして、デバイスの解像度(画面サイズ)に応じてレイアウトを自動的に変更するレスポンシブウェブデザインがある。

レスポンシブウェブデザインは、PC、タブレット、スマートフォンの画面サイズを基準にデバイスの種類を判断し、CSS3 で追加された Media Queries[88]の機能を利用し、1 つの HTML データをデバイスごとに異なるレイアウトやデザインで表示する手法である。

従来はデバイスごとに HTML を用意する方法が取られていたが、メンテナンスの手間が増え、URL をデバイスごとに分けなければならなかった。レスポンシブウェブデザインにより、複数のデバイスでも同一の HTML で適切な表示ができるため、メンテナンス性が向上する。

デジタルアーカイブ・システムでは、国立公文書館デジタルアーカイブにおいて採用されている。

---

[87] 株式会社 ICT 総研『2016 年度 スマートデバイス市場動向調査』、2016 年 7 月 http://ictr.co.jp/report/20160706.html （2017 年 1 月 24 日アクセス）

[88] http://www.w3.org/TR/css3-mediaqueries/ （2017 年 1 月 24 日アクセス）

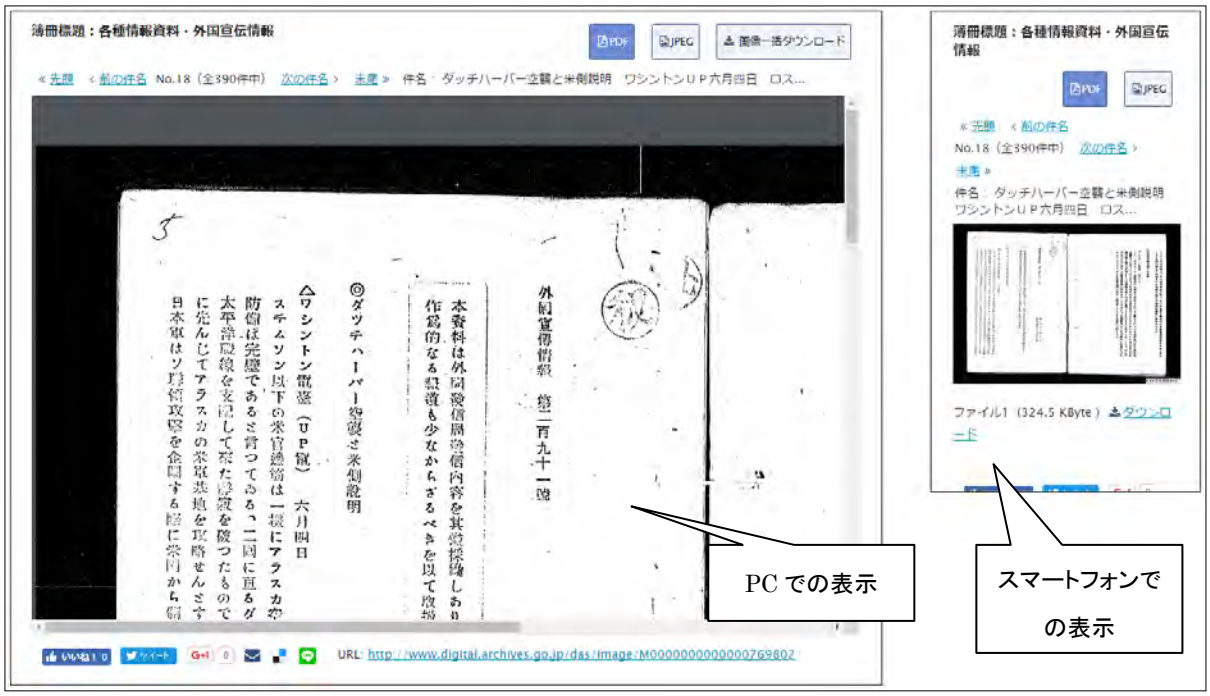


図 30 国立公文書館デジタルアーカイブの画像閲覧における、PC とスマートフォンの表示例

(3) ビューア

「2.2.9 デジタルコンテンツ配信技術」の「デジタルコンテンツフォーマット」で配信フォーマットとして挙げた JPEG、MP4、MP3 は、OS を問わず標準搭載のソフトウェアで閲覧することができる。

また、デジタルアーカイブ・システムにおいて画像を配信するためのフレームワークである IIIF の採用が広まりつつある。配信する側が IIIF に対応することで、利用者はビューアを自由に選択することが可能となる。

IIIF に対応したビューアである OpenSeadragon[89]は、JavaScript で実装されたオープンソースの画像ビューアである。様々な解像度の画像を、解像度ごとにタイル画像に分割しておくことで、スムーズな拡大・縮小と移動を行える操作性が特徴である。ただし、この機能を構築するには、タイル画像を事前に用意する必要があり、元の画像データの数倍から数十倍のディスク容量が必要となる。

(4) ネットワーク

一般利用者が利用する回線は、「2.2.3 ネットワーク環境」の調査結果と同様に光回線が増えていると考えられるが、スマートフォンやタブレットといったスマートデバイスが普及し、高速モバイル通信の利用が増加している。特に下り通信速度が 75Mbps から 100Mbps の LTE(Long Term Evolution)[90]の加入者数は 2012 年以降急激に増加しており、スマートデバイスでも高速通信が普及している。ただし、LTE では利用者数の増加に伴いトラフィック圧迫の懸念が高まり、通信キャリア各社は、例えば月当たりのデータ通信量の上限を設けるといった対応をとっている。上限に達すると通信速度が 128Kbps に制限され、データ通信の利便性が大きく損なわれてしまう。大容量データの配信ではデータ量の圧縮等、利用者への配慮

---

89 http://openseadragon.github.io/ (2017 年 1 月 24 日アクセス)

90 第 3 世代携帯電話(3G)と第 4 世代携帯電話(4G)の中間技術として 3.9G と呼ばれることもある。

が必要である。

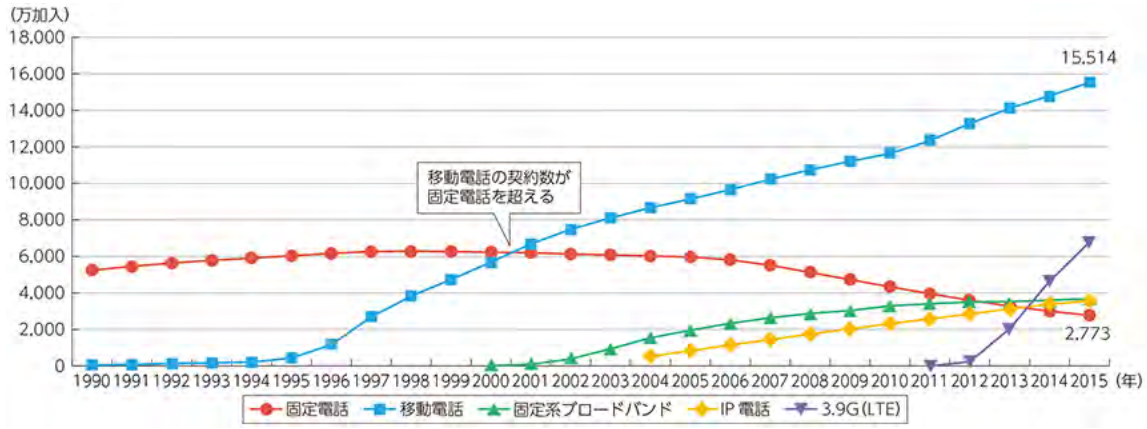


図 31 通信サービス加入契約者数の推移[91]

(5) アクセシビリティ

幅広い利用者がデジタルアーカイブ・システムを利用できるよう、ウェブアクセシビリティを取り巻く状況について調査する。

(ア) ウェブアクセシビリティの標準化動向

ウェブアクセシビリティとは、バリアフリーなウェブコンテンツを作成する方法を提示し、障害のある人がインターネットのウェブへ容易にアクセスできるようにすることを目的とした取組である。

これまでウェブアクセシビリティにおける基準の標準化が進められており、国内においては、2004 年 6 月に「高齢者・障害者等配慮設計指針 情報通信における機器，ソフトウェア及びサービス」(情報アクセシビリティ JIS)の個別規格として、「第 3 部:ウェブコンテンツ」(JIS X 8341-3:2004)が公示された。また、国際的には、W3C(WWW コンソーシアム)が、ウェブコンテンツ アクセシビリティ ガイドライン (WCAG) 1.0 を 1999 年に勧告、2012 年には同ガイドラインを改正した、ウェブコンテンツ アクセシビリティ ガイドライン (WCAG) 2.0 が ISO による国際標準「ISO/IEC 40500:2012(情報技術 -- W3C ウェブコンテンツ アクセシビリティ ガイドライン (WCAG) 2.0)」として規格化された。その後、2016 年 3 月に、国内標準 JIS X 8341-3:2004 は対応する国際規格 ISO/IEC 40500:2012 との完全な一致を図るため、JIS X 8341-3:2016 として改正されている。[92]

(イ) 自治体におけるウェブアクセシビリティの動向

総務省では、高齢者や障害者を含む誰もが公共分野のホームページやウェブシステムを利用することができるよう、平成 16 年 11 月から「公共分野におけるアクセシビリティの確保に関する研究会」を開催し、

---

91 総務省『情報通信白書』 平成 27 年版
http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h27/html/nc111220.html (2017 年 1 月 24 日アクセス)

92 http://www.soumu.go.jp/main_sosiki/joho_tsusin/b_free/b_free02.html (2017 年 1 月 24 日アクセス)

地方公共団体で活用できる運用モデルについて検討を進め、平成 17 年 12 月に本研究会で策定された、地方公共団体等におけるウェブアクセシビリティの確保・向上のための取組モデルである「みんなの公共サイト運用モデル」を公表した。

平成 22 年の改訂を経て、平成 28 年に障害者差別解消法が施行されたこと及び JIS X 8341-3 が改正されたこと等を踏まえて、平成 28 年 4 月に運用モデルを改定し、「みんなの公共サイト運用ガイドライン（2016 年版）[93]」が公表された。[94]

「みんなの公共サイト運用ガイドライン」は、国及び地方公共団体等公的機関の「ウェブアクセシビリティ」対応を支援するために総務省が作成したガイドラインであり、実施すべき取組項目と手順、重視すべき考え方等を解説している。この中では、各団体として取り組む対象範囲、適合レベルは各団体がガイドラインとして取りまとめることとなっている。また、スマートフォンからのアクセスへの対応も明記されている。

(6) まとめ

近年主要なブラウザは HTML5 や CSS3/4 に対応しており、画像や動画・音声データを取り扱うことが容易になったことから、ブラウザのみで高機能なユーザーインターフェースが実現可能となってきた。なお、このような状況においても、ブラウザの種類やバージョンに依存する機能がいまだ実装されていることから、利用者への提供情報を設計する際には注意を要する。

スマートデバイス、とりわけスマートフォンの普及率は増加しており、デジタルアーカイブ・システムをスマートデバイスから閲覧する利用者も増加すると予想される。利用者拡大のためにはマルチデバイスに対応したユーザーインターフェースが必要である。多面的なデザインへの対応からその分のコストを必要とするが、マルチデバイスに対応したユーザーインターフェースを実現する手段として、レスポンシブウェブデザインを採用することが考えられる。

デジタルコンテンツの閲覧環境としては、ブラウザもしくは OS に標準装備される環境が整備された結果、新たなアプリケーションソフトウェアの導入を要せずにデジタルコンテンツの閲覧が、以前と比べ、多くの Web サイトにおいて可能となってきている。また、画像コンテンツにおいては、コンテンツの配信側が IIIF に対応するとともに、IIIF に対応した画像ビューアが普及することにより、利用者の利便性向上や画像コンテンツの相互運用性の向上が期待されるところにある。

利用者におけるネットワーク接続環境としては、ネットワークケーブルによる固定的な通信回線だけではなく、無線通信技術を用いた高速モバイル通信の普及がスマートデバイスの普及を後押ししており、このスマートデバイスの利用者（加入者）増はいまだ伸び悩むところまでは至っていない。しかし、高速モバイル通信においては、利用者増に伴う、トラフィック圧迫の懸念から、例えば、利用状況に応じて、月当たりのデータ通信量の上限を設ける等利用を制限するといった通信キャリアがある。

ウェブアクセシビリティの面においては、バリアフリーなウェブコンテンツの作成を目指した、ウェブアクセシビリティに関する基準が国内外において規格化されている。また、公的機関においては、総務省がウェブアクセシビリティの確保・向上を目的とした、公共サイトの運用に関するガイドラインを公表しており、ウェブ

---

93 http://www.soumu.go.jp/main_content/000439213.pdf （2017 年 1 月 24 日アクセス）
94 http://www.soumu.go.jp/main_sosiki/joho_tsusin/b_free/b_free02.html （2017 年 1 月 24 日アクセス）

アクセシビリティを意識したウェブサイトを構築するための環境が整備されてきている。公的機関が運営するデジタルアーカイブ・システムでは、各機関でウェブアクセシビリティガイドラインを定め、それらの機関が構築するウェブサイトにおいて、このガイドラインに準拠することが求められる。

2.3. デジタルアーカイブ・システムに活用可能と考えうる技術に係る調査

2.3.1. OCR 技術と画像検索技術

資料画像から文字情報を機械的に判読してデータ化することにより、メタデータを充実させ、デジタルアーカイブ・システム利用者の検索における利便性向上を図ることが可能になると考えられる。本項では、画像からの文字読み取りや解析等に係る最新動向について調査する。

(1) OCR 技術

OCR（Optical Character Recognition/Reader）とは、スキャナやカメラ等で文字を読み取る技術である。OCR の基本的な流れは、次の通りである。

- 読み取った画像に対し、表題や段組みといったレイアウト解析を行う。
- レイアウト解析の結果から、行を特定して切り出し、更に 1 文字ずつ切り出す。
- 一定の大きさに変換(正規化)し、上下、左右、斜め方向に分解(特徴抽出)する。
- 標準パターン[95]と正規化し特徴抽出したものとのマッチングを行う。
- 「タ（漢字）」と「タ（カタカナ）」や「カ（漢字）」と「カ（カタカナ）」などを認識するために前後関係を見て判断(知識処理)する。

マッチングや知識処理の改善により、文字の認識率が向上している。現代において作成、印刷、流通する活字文書に対しては、高解像度でのスキャニング処理を施すことにより、99%を超える解読正解率を得られることが複数の OCR ベンダーより報告されている。

公文書館の所蔵資料における全文テキストデータの取得は、取得データを基にした目録データの整備に係る作業負担の軽減、多様な情報をデータとして取得し、検索対象データとして処理することによる検索精度の向上等への活用に有用なものとなると考えられる。

(2) 画像検索技術

画像検索技術とは、類似した他の画像を検索する技術である。大量の画像データの中から目的の画像を素早く探し出すために、色や模様といった表面的な類似性ではなく、映っているオブジェクト（時計や花などの物体）を考慮した検索を行う。

与えられた画像の種（クエリ画像）からそのオブジェクトの「見た目の特徴」を学習し、その学習結果を利用して類似したオブジェクトが映る画像を検索する。例えば、「飛行機」が映っている画像をクエリー画像として与えることで、「飛行機」が映る他の画像を検索することができる。複数のクエリー画像を与えることができ、撮影アングルなどが異なる画像を追加することで検索精度を向上させることが可能である。

---

[95] あらかじめ登録しておいた認識したい文字を様々なフォント(字体)や「かすれ」、「つぶれ」文字の認識を安定して行うために、いろいろな状態で印字された文字を平均化して作る。

図 32 画像検索における学習の仕組みと画像検索の例[96]

(3) 最新動向

OCR 技術と画像検索技術を組み合わせた技術として、「くずし字検索」がある。

「くずし字」を楷書に変換する翻刻は、これまで全て専門家が手作業で行っていたが、ソフトウェアによる支援で作業の一部を自動化し、翻刻の高速化、低コスト化を実現できるようになった。画像検索技術と、従来の OCR 技術を組み合わせることで、特定の条件下では精度 80%以上で文字の識別が可能であることが実証されている[97]。国文学研究資料館の「日本語の歴史的典籍の国際共同研究ネットワーク構築計画」ではくずし字 OCR を適用した実証試験が実施されている。

---

96 http://www.nes.or.jp/transfer/catalog/2016/01/50a/ (2017 年 1 月 24 日アクセス)

97 山本 純子、大澤 留次郎「古典籍翻刻の省力化 くずし字を含む新方式 OCR 技術の開発」『情報管理』2016 年、vol. 58、no. 11、p819-827
https://www.jstage.jst.go.jp/article/johokanri/58/11/58_819/_pdf (2017 年 1 月 24 日アクセス)

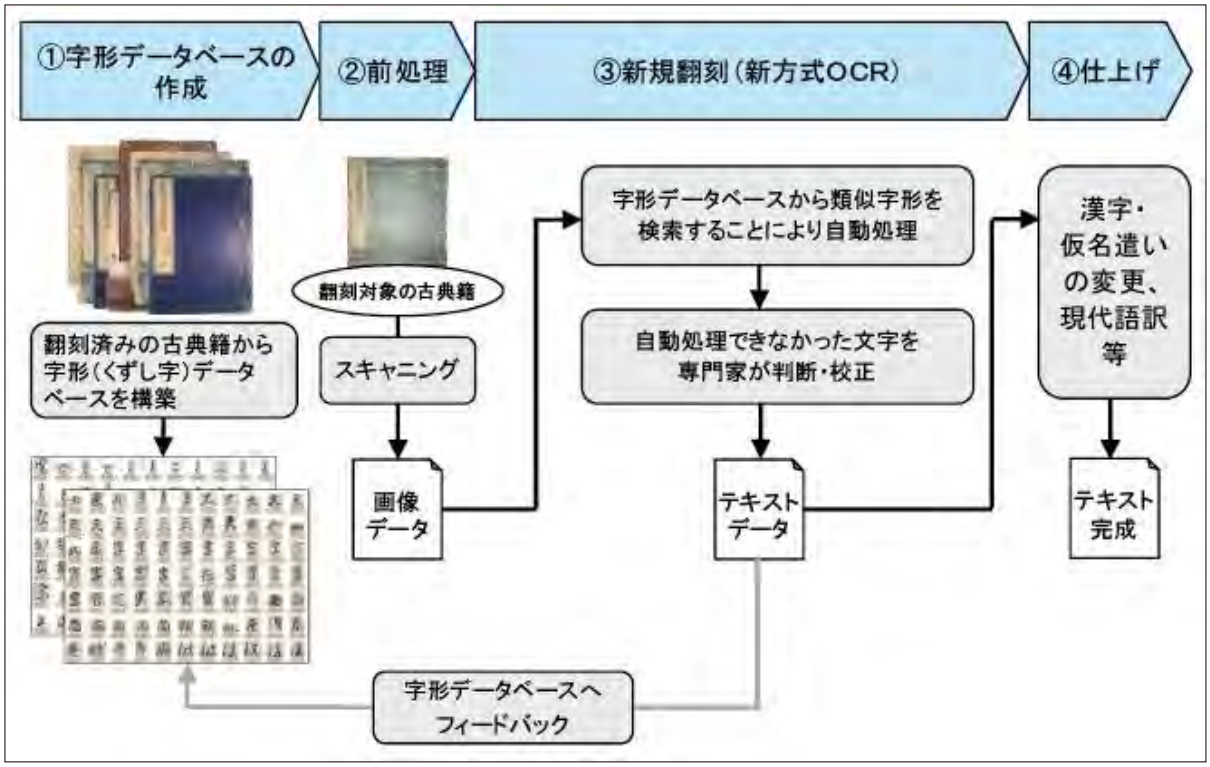


図 33 新方式くずし字 OCR を用いた翻刻の工程[98]

(4) まとめ

OCR の文字読取り精度は向上しており、高い解読正解率を取得可能な文書もあるが、このような文書は、公文書館が所蔵する資料からすると一部の文書に止まるものであり、より古い年代に作成された文書に対しては、活字文書であっても、文書の紙質や印刷の状態によっては解読正解率が下落するものと推測される。公文書館の所蔵資料への OCR 技術の適用可能性については、処理対象資料に対する認識試験を重ね、例えば全文テキストデータの取得計画を作案の上、判断することになると考える。また、OCR 技術と画像検索技術と組み合わせたくずし字検索は、識字における 80%の精度も特定条件下における実証試験段階のものであり、公文書等にみられる手書き文字、古文書におけるくずし字等への OCR 技術の活用等への実用化には、まだ時間がかかると推測される。今後の研究の進展やかかる技術の発展に注視すべきと考える。

---

98 蒲生秀典「日本の古典籍・歴史資料のデジタル化における新潮流」文部科学省科学技術・学術政策研究所『STI Horizon』Vol.2、No.3、2016 年 9 月
http://www.nistep.go.jp/activities/sti-horizon%E8%AA%8C/vol-02no-03/stih00039（2017 年 1 月 24 日アクセス）

2.3.2. 自動翻訳

資料の利用促進において、日本語から複数の言語へ、また複数の言語から日本語への自動翻訳が有用であると考えられる。本項では、自動翻訳に関する技術について最新動向を調査する。

(1) 自動翻訳技術

機械翻訳では、従来は文法をもとに翻訳していく手法が取られていた。言語データの統計モデル化が可能となり、対訳データに基づく翻訳手法が急速に進展している。データの蓄積とハードウェア性能の向上により、翻訳精度が向上している。

(2) 自動翻訳の事例

自動翻訳の事例として、「言語グリッド」がある。言語グリッドは、集合知のアプローチにより、世界中の言語資源(辞書、機械翻訳など)を共有することができる、インターネット上の多言語サービス基盤である。言語グリッドを使うことで、インターネット上の言語資源(対訳辞書など)や言語処理機能(機械翻訳など)を自由に組み合わせて使ったり、コミュニティが作った言語資源を追加し、コミュニティの活動に特化したWeb サイトを作ったりすることができる。

図 34 言語グリッドの活用事例[99]

(3) まとめ

自動翻訳技術の研究は進められているが、デジタルアーカイブ・システムでは公文書や歴史的資料に特有の表現があり、対訳辞書等の情報整備が必要と考えられる。

---

99 石田亨「サービスコンピューティングに基づく多言語サービス基盤の実現」(2012 年 10 月) http://www.soumu.go.jp/main_sosiki/joho_tsusin/scope/event/h24yokousyu/B1-2(pre).pdf (2017 年 1 月 24 日アクセス)

2.3.3. 辞書/シソーラス検索機能と利用可能な辞書データ

デジタルアーカイブ・システムにおける検索機能の向上の手段として、辞書やシソーラスを使った検索技術について最新動向を調査する。

(1) 辞書/シソーラス検索技術

シソーラスは、語句を意味の関係から分類した辞書やデータベースのことである。語句の意味によって、大分類から小分類へと体系的に分類されており、同義語や上位概念、下位概念を意味する語句を木構造や表形式で表現している。キーワードを用いた検索の際、そのままのキーワードでのみ検索すると、同じ意味でも表現が異なる場合には条件に一致せず抽出されない。検索システムにシソーラスを用いることで、このような検索漏れを防止することができる。

検索システムで利用できる辞書、シソーラスには様々なものがある。有償のものとして、概念に関する情報を記載した辞書、日本語単語を分類したシソーラス、機械翻訳システム ALT-J/E[100]のために開発された日本語シソーラスがある。

無償のものとして、伝染病関連の語に関するオントロジー[101]、Princeton WordNet 3.0[102]の類義関係のセットに対して日本語を付与したもの、自然言語処理用の動詞辞書がある。

(2) シソーラス活用の最新事例

シソーラスの活用事例には、科学技術振興機構(JST)が運営するJ-GLOBALがある。研究者情報、文献情報、特許情報、研究課題情報、機関情報、科学技術用語情報、化学物質情報、資料情報等の総合的学術情報データベースである。 全ての情報が無料で公開されており、オリジナル文献を入手するためのリンクもある。様々な種別のデータベースを統合することによって、例えば文献と特許を人(著者・発明者)で接続する等、研究開発でキーとなる情報同士の連結を実現している。また、着目した情報について連携する外部サイトリンクが付与されていることによって、関連情報の調査を支援している[103]。

また、アジア歴史資料センターの資料検索システムでは、同義語や関連語、表記ゆれを辞書としてデータベース化し、検索において利用している。

(3) まとめ

デジタルアーカイブ・システムでも辞書検索を備えた事例はあるが、公文書や古書・古文書といった資料を効果的に検索できる辞書データは一部の機関でしか整備されていない状況である。これらの資料の特徴を踏まえた辞書やシソーラスの整備が課題であるが、整備にかかるコストを抑えるためにも、各機関の持つ辞書データを共通化し、相互利用できる仕組みが期待される。

---

100 開発中の日英翻訳システム。

101 対象世界にかかわる諸概念を整理して体系づけ、コンピュータにも理解可能な形式で明示的に記述したもの。

102 プリンストン大学で開発された英語の大語彙のデータベース。

103 https://jglobal.jst.go.jp/footer/?page=aboutus (2017 年 1 月 24 日アクセス)

2.4. デジタルアーカイブ・システムへの適用に係る検討

本調査で調査検討した各種システム構成や、各種 IT 技術について、「活用の可能性」「稼働環境」「求められる技術要件」「コスト」「課題」の観点により、デジタルアーカイブ・システムへの適用について検討する。

2.4.1. システム構成等のデジタルアーカイブ・システムへの適用

(1) 単館運用

(ア) オンプレミス型

活用の可能性	構築や運用において最も柔軟性のある構成であり、活用可能である。
稼働環境	電源や空調が整備されたサーバ室や、ネットワーク回線等が自館に必要である。
求められる技術要件	システム基盤とアプリケーションに求められる機能要件の取りまとめが必要である。構築後にシステム基盤とアプリケーションの運用管理を担当する職員が必要である。
コスト	システム基盤の初期コストが高くなる。アプリケーションは、開発・導入型よりも、パッケージ導入型の方がコストは抑えられる。
課題	専門知識を持った担当者が必要。 施設改修やハードウェアの購入など、初期費用がかかる。

(イ) コロケーション型

活用の可能性	構築や運用において柔軟性を保ちながら、安定した環境で 24 時間稼働させることに適した構成であり、活用可能である。
稼働環境	データセンターにハウジングする。
求められる技術要件	システム基盤とアプリケーションに求められる機能要件の取りまとめが必要である。
コスト	システム基盤の初期コストが高くなる。データセンターの利用にかかる費用が必要だが、ネットワーク回線や、サーバ室を整備するコスト等が抑えられる。 アプリケーションは、開発・導入型よりも、パッケージ導入型の方がコストは抑えられる。
課題	専門知識を持った担当者が必要。ハードウェアの購入など、初期費用がかかる。

(2) 共同運用

(ア) プライベートクラウド型(開発・導入型／パッケージ導入型)

活用の可能性	専用のクラウド環境でリソースを効率的に活用しながら、柔軟性のあるアプリケーションを導入できる構成であり、活用可能である。
稼働環境	クラウド事業者が提供する専用のクラウド環境、または共同運用機関が構築する専用のクラウド環境となる。
求められる技術要件	アプリケーションに求められる機能要件の取りまとめが必要である。 アプリケーションの運用管理を担当する職員が必要である。
コスト	システム基盤の初期コストは抑えられ、月額の運用コストが発生する。

課題	システム基盤の初期コストと運用負担が抑えられるが、パブリッククラウドほどコストメリットは出ない。

（イ）プライベートクラウド型（サービス利用型）

活用の可能性	提供されるサービスが各機関の要求に一致する場合には活用可能である。 ただし、サービスを提供する機関は限られることが想定され、活用できるケースは限られると考えられる。
稼働環境	クラウド事業者が提供する専用のクラウド環境、または共同運用機関が構築する専用のクラウド環境となる。
求められる技術要件	提供されるサービスに適合するデータの用意が必要である。
コスト	インフラ、アプリケーションとも初期コストは抑えられ、月額の運用コストが発生する。
課題	パブリッククラウドほどコストメリットが出ない。

（ウ）パブリッククラウド型（開発・導入型／パッケージ導入型）

活用の可能性	クラウド環境で初期投資を抑えながら、柔軟性のあるアプリケーションを導入できる構成であり、活用可能である。 ただし、クラウド環境の制約（セキュリティ等と取り扱うデータの種類等）について検討が必要であり、活用できるケースは限られると考えられる。
稼働環境	クラウド事業者が提供するクラウド環境となる。
求められる技術要件	アプリケーションに求められる機能要件の取りまとめが必要である。 アプリケーションの運用管理を担当する職員が必要である。
コスト	システム基盤の初期コストは抑えられ、月額の運用コストが発生する。
課題	システム基盤の初期コストと運用負担が抑えられるが、ネットワークセキュリティ等、提供されるサービス内容の制約について検討が必要である。

（エ）パブリッククラウド型（サービス利用型）

活用の可能性	提供されるサービスが各機関の要求と一致する場合には活用可能である。
稼働環境	クラウド事業者が提供するクラウド環境となる。
求められる技術要件	提供されるサービスに適合するデータの用意が必要である。
コスト	システム基盤、アプリケーションとも初期コストは抑えられ、月額の運用コストが発生する。
課題	コスト面では最も抑えられるが、業務とサービスの適合度、ネットワークセキュリティ等、提供されるサービス内容の制約について検討が必要である。

2.4.2. クラウド等 IT 技術のデジタルアーカイブ・システムへの適用

(1) クラウド技術

活用の可能性	短期間にインフラを構築する場合や、複数の機関で共同運用する場合に活用できる。
稼働環境	クラウド技術を活用したシステム構成には（クラウドサービスの提供形態）として、IaaS, PaaS, SaaS がある。ニーズに合わせた柔軟な構成を採用することが可能である。
求められる技術要件	仮想化、分散化、自動化にかかる技術等、利用するクラウドサービスの形態に応じて必要となる技術要件は異なる。
コスト	オープンソースで提供されているクラウドインフラストラクチャソフトウェアもあり、安価にインフラを構築できる。また、共同運用する場合は複数の機関による運用コストの分担も考慮可能である。
課題	採用を検討することになったクラウド技術を活用したシステム基盤が、デジタルアーカイブ・システムを構築する際の要求要件に合うサービスや機能を提供することが可能であるか見極める必要がある。

(2) サーバ環境

（ア）サーバ技術(仮想化技術含む)

活用の可能性	サーバ、OS、ミドルウェアは、高機能化や低価格化、オープンソース化が進んでおり、導入に当たっての選択肢は増えている。 また、仮想化技術は技術の発展及び普及が進んでおり、限られたサーバ等のハードウェア資源を有効に活用するための設計における方策の一つとして、考慮することができるものと考えられる。
稼働環境	デジタルアーカイブ・システムを実現するためのサーバ環境は年々高機能化が進んでおり、多様な稼働環境のなかから、ニーズにあった環境を選択可能な状況にある。
コスト	サーバ機器の性能向上に伴い、十分な性能と安定性を確保できる環境が低コストで導入可能となってきている。 また、サーバ環境において仮想化技術を採用することにより、ハードウェア資源を有効に活用し、構築コストを抑えることが可能である。

（イ）ストレージ技術

活用の可能性	ストレージ技術において、その基礎となる技術は普及してから相当の年月が立っており、安定した技術として提供されている。また、近年、クラウド技術を活用したクラウドストレージが普及してきている。クラウドストレージでは、オンプレミス等の形態で扱うストレージ装置と比べ、提供可能なサービスが限定されるが、必要とするディスク容量に絞って、サービスを受けること等が可能である。
コスト	ストレージの容量当たりの単価は減少傾向にある。また、クラウドストレージにおいては、利用するディスク容量を必要な容量に絞ることでコストを抑えることが可

	能である。
課題	大容量のデータを扱うためのコスト面でのハードルは低くなってきているものの、クラウドストレージ等導入する形態によっては、実現可能なサービスや機能が制限される。

(3)　ネットワーク環境

活用の可能性	インターネット接続回線には、公開するコンテンツ、想定のアクセス数、予算に応じた適切な回線を用意する必要がある。現時点では様々なネットワークプロバイダが多様なニーズに合わせた回線を提供しており、構築するデジタルアーカイブ・システムに合わせた回線の選択することが可能である。デジタルアーカイブ・システム等 Web ベースでサービスを提供するシステムにおいては、通信プロトコルに HTTP を用いることが前提となるが、同プロトコルにおいて、HTTP/2 を採用することにより利用者における閲覧速度の向上が期待できる。 また、デジタルアーカイブ・システムで非公開データを扱う場合には、ネットワーク暗号化技術（HTTPS）は必須要件である。HTTPS を採用することにより、デジタルアーカイブ・システムと利用者の間で交換する情報に係るセキュリティを確保することが可能である。
コスト	既存のインターネット回線があれば、デジタルアーカイブ・システムの接続に必要な要件を満たしている場合、その回線を使うことで、コスト面における合理化を図ることができるものと考えられる。

(4)　バックアップ関連技術

活用の可能性	データ毎の重要性や更新頻度を勘案の上、それぞれのデータ特性に合ったバックアップ技術、バックアップ方式、バックアップメディアの組み合わせが可能である。
コスト	保有するデータ量と適切なバックアップ方式やメディア等を検討することにより、コストを抑えることが可能である。
課題	効率的なバックアップ運用を実現するためには、各機関の保有するデータ量やデータの増加予測等に基づく検討を要する。

(5)　ネットワークセキュリティ技術

活用の可能性	コンピュータネットワークにおける脅威は年々増大し、デジタルアーカイブ・システムにおいても安定的なサービスの提供、データ保全のためにセキュリティ対策を講じる必要がある。最低限の対策としてウイルス対策ソフトと WAF の導入が推奨される。
求められる技術要件	導入の難易度はセキュリティレベルにより異なる。
コスト	セキュリティレベルに比例してコストが高くなる傾向にある。
課題	ネットワークセキュリティに係る技術等の導入コストはセキュリティレベルに応じて高くなる傾向にある。各機関のセキュリティ対策方針等に応じて適切な対応を選

	ぶ必要がある。また、IPA（独立行政法人情報処理推進機構）等の公表資料で示されるセキュリティ対策も有効な手段である。

(6)　目録関連技術

（ア）EAD

活用の可能性	アーカイブズにおける符号化記録史料記述の事実上の国際標準であるEADに対応することで国内外を問わず目録情報の交換の容易さや機械可読性が高まり、目録情報の利活用を拡大させる効果も期待できる。
課題	EAD の最新版である EAD3 のデジタルアーカイブ・システムへの実装例を本調査の時点では確認することができなかった。EAD3 は 2015 年に公表されており、システムのライフサイクルを勘案すると、現時点ではそれほど多くの時間が経過しているわけではない。今後の普及が見込まれると考えられる。

（イ）Dublin Core

活用の可能性	横断検索に係るデータ項目として Dublin Core を採用することは、システム間連携において有用である。DCMES は簡潔であることから様々なシステム間の連携を円滑に実現することが期待できる。
課題	DCMES は簡潔であることから多様な領域に適用できる一方、表現可能な情報の粒度が粗く、相互運用性を低下させる可能性がある点が課題である。DC タームス等の Dublin Core の拡張項目を用いる等、システム間で共通する、より多くのデータ項目を扱うことで、検索の精度をより高めることが可能である。また、アプリケーションプロファイルを適用することで相互運用性を高めることが可能である。

（ウ）パーマリンク

活用の可能性	資料目録やデジタルコンテンツの情報へのアクセスを永続的に保証するパーマリンクは、デジタルアーカイブ・システムにおいて、長期安定利用のために非常に重要な機能である。
求められる技術要件	リソースを一意に識別するための識別子の設計が必要である。このような識別子の実現に当たっては Cool URI や DOI の活用が挙げられる。
課題	デジタルアーカイブ・システムにおいて、提供する情報等へのアクセス方法が変わらないことの保証は重要な機能である。 また、システムリプレースは、デジタルアーカイブ・システムで提供するコンテンツを表す識別子が変更されるきっかけとなりうるものである。このような識別子の変更を避け、URI が維持されることを明示的に仕様として入れることにより、パーマリンクを継続する等の対応が考えられる。

(7)　目録データの提供

（ア）目録データの提供

活用の可能性	デジタルアーカイブ・システムが有する目録データは、利用者が所蔵資料にアク

	セスするために用いるものであることは当然であるが、他のシステムが提供するサービスで活用可能とすることにより、目録データそのものの利活用の機会を拡大することが可能である。
稼働環境	RDF に対応する場合、RDF ストアや SPARQL に対応したソフトウェアが稼働する環境である必要がある。
コスト	RDF を用いてデータを他のシステムへ提供するための実装にかかるコストが発生。
課題	デジタルアーカイブ・システムでは、Web ブラウザ上で目録を閲覧可能とすることに加え、CSV 等の形式でデータを一括ダウンロードできることが望ましい。また、提供するデータを他のシステムによるサービスが活用可能とするために、RDF 等の機械可読性の高い形式によるデータ提供を行えることが、データの相互運用性向上という点から望ましい。

（イ）オープンデータ

活用の可能性	国のオープンデータ戦略が進められている等、デジタルアーカイブ・システムで提供するデータは、他のシステムが活用可能となるようオープンデータへ対応することにより、より広範なデータの提供が求められている。
求められる技術要件	オープンデータを実現する技術には、RDF や LOD、WebAPI がある。
課題	オープンデータの推進にあたり、技術的にはデジタルアーカイブ・システムへの WebAPI の実装が考えられる。また、デジタルアーカイブ・システムで提供するデータの利用規約として、政府標準利用規約やクリエイティブ・コモンズといったライセンスに関する標準を踏まえつつ、各機関の事情を考慮して判断することが求められる。加えて、オープンデータの推進には、アーカイブズ所蔵機関が固有の識別子を定め、維持することが望まれる。

(8)　情報検索技術

（ア）検索エンジン技術

活用の可能性	デジタルアーカイブ・システムでは、目録情報に非構造化データを扱うことも想定される。そのための対応として、目録検索における全文検索に対応した技術（全文検索エンジン）の採用が考えられる。全文検索エンジンは、高機能・高性能なものがオープンソースで提供されており、導入のハードルは決して高いものではなくなってきている。また、全文検索機能を有する RDBMS を用いることにより、データ管理の効率化を図る方法も考えられる。
稼働環境	広く普及している技術であり、稼働環境を選ばない。
コスト	オープンソースのソフトウェアを利用することでコストを抑えられる。

（イ）横断検索技術

活用の可能性	デジタルアーカイブ・システムでは、HTTP の利用を前提としたウェブベースの検索プロトコルである SRU/W、OpenSearch のいずれかを実装することが、相互運用性の向上という点からも推奨される。

稼働環境	SRU/W、OpenSearch のいずれのプロトコルも特別な稼働環境を要するものではない。
コスト	各機関で導入するパッケージやサービスが上記プロトコルに対応していない場合は、サーバ機能等の実装にかかるコストが必要となる。

(9) デジタルコンテンツ配信技術

（ア）デジタルコンテンツの配信

活用の可能性	ストリーミング方式は、ファイルサイズが大きい動画や音声を配信する際に活用可能である。プログレッシブダウンロードによる配信も選択肢としてはありうる。
稼働環境	仕組みにより異なる。ストリーミング方式を採用する場合は、この方式に対応したコンテンツ配信のためのサーバソフトウェアの導入を要することとなる。
コスト	ストリーミング方式は配信用ソフトウェア（有償）が必要である。 ダウンロード方式、プログレッシブダウンロード方式ではデータ提供のために必要となるソフトウェアはない。

（イ）デジタルコンテンツの配信に関わる有益な技術

活用の可能性	画像配信の国際的フレームワークである IIIF は画像データの相互運用性向上の観点からもデジタルアーカイブ・システムへの活用が可能であると考える。 コンテンツキャッシュ技術は、サーバの負荷軽減や表示速度の向上といった特徴があるが、費用対効果の面から、想定されるアクセス数に応じて導入を検討する必要があると思われる。
コスト	IIIF への対応にはコストがかかる。 コンテンツキャッシュ技術は有償のサービスが多く、導入にはコストがかかる。
課題	IIIF については配信するコンテンツに応じてデジタルアーカイブ・システムでの採用を判断すべきである。 コンテンツキャッシュ技術はアクセス数が少ない場合、費用対効果は低いと考えられるため、想定される利用者(アクセス)数と予算に応じて、導入を検討する必要がある。

(10) クライアント環境

（ア）ブラウザ等ソフトウェア技術

活用の可能性	主要なブラウザは HTML5、CSS3/4 への対応が進んでおり、ブラウザのみで操作性の高いユーザーインターフェースや機能を実現することが可能である。
求められる技術要件	HTML5、CSS3/4、JavaScript といった技術が必要である。
稼働環境	HTML5、CSS3/4 に対応したブラウザである必要がある。
コスト	広く普及している技術であるため、低コストでの開発が可能である。
課題	ブラウザのバージョンアップサイクルは早く、実装状況が変動することがあるため動向を注視する必要がある。

（イ） PC 等ハードウェアへの対応技術(スマートデバイスへの対応を含む)

活用の可能性	レスポンシブウェブデザインは普及が進みつつあり、多様なデバイスに適したユーザーインターフェースが可能である。
求められる技術要件	CSS3 の Media Queries 技術。
コスト	レスポンシブウェブデザインへの対応にコストがかかる。
課題	スマートデバイスの普及は著しく、これらによる利用が今後も増加する傾向にあると思われる。レスポンシブウェブデザイン等による、マルチデバイスに対応したユーザーインターフェースの提供を可能とすることが望ましい。

（ウ） ビューア

活用の可能性	JPEG、MP4、MP3 は OS を問わず標準搭載のソフトウェアで閲覧することができる。 配信する側が IIIF に対応することで利用者はビューアを自由に選択することが可能となり、利用者の利便性向上やシステム間での画像データ連携も容易に行えることが期待できる。

（エ） ネットワーク

活用の可能性	ネットワークケーブルによる固定的な通信回線だけではなく、無線通信技術を用いた高速モバイル通信の普及がスマートデバイスの普及を後押ししており、このスマートデバイスの利用者（加入者）増はいまだ伸び悩むところまでは至っていない。
課題	高速モバイル通信においては、利用者増に伴う、トラフィック圧迫の懸念から、例えば、利用状況に応じて、月当たりのデータ通信量の上限を設ける等利用を制限するといった通信キャリアがある。 このような対応は利用者における利便性を損なうことから、利用者の利便性に配慮した対策や新たな技術の出現が待たれる。

（オ） アクセシビリティ

活用の可能性	ウェブアクセシビリティを考慮した、バリアフリーなウェブコンテンツの作成を目指し、そのための方法が国内外において規格化されている。また、公的機関においては、総務省がウェブアクセシビリティの確保・向上を目的とした、公共サイトの運用に関するガイドラインを公表しており、ウェブアクセシビリティを意識したウェブサイトを構築するための環境が整備されてきている。
課題	総務省による上記ガイドラインでは、同ガイドラインを踏まえた、ウェブアクセシビリティ方針の策定と公開を各機関が実施することとしている。ウェブアクセシビリティにかかる様々な対応の実現が求められている。

2.4.3. デジタルアーカイブ・システムに活用可能と考えうる技術

(1) OCR 技術と画像検索技術

活用の可能性	くずし字検索は実証試験段階であり、限定された条件下においてではあるが、80%に及ぶ精度で文字を認識するといった成果を挙げている。実用化にはまだ時間がかかると推測されるが、今後の進展が期待される。
稼働環境	OCR は市販ソフトウェアも多く、多様な稼働環境に対応している。 画像検索技術は研究段階であり実用化の途上である。
求められる技術要件	OCR の文字読取り技術、画像認識技術
コスト	くずし字の字形データセットがオープンデータとして公開される等の動向はあるが、現時点では実証試験段階であり、導入コストは明確でない。

(2) 自動翻訳

活用の可能性	自動翻訳技術の研究は進められているが、公文書館の所蔵資料への適用を考慮した場合、公文書や歴史的資料に特有の表現に関する、対訳辞書等の情報整備が必要と考えられる。

(3) 辞書/シソーラス検索機能と利用可能な辞書データ

活用の可能性	デジタルアーカイブ・システムでも辞書検索を備えた事例はあるが、公文書や古書・古文書といった資料を効果的に検索できる辞書データは一部の機関でしか整備されていない状況である。
コスト	辞書データの整備にコストがかかる。
課題	デジタルアーカイブ・システムで扱う、公文書館の所蔵資料の特徴を踏まえた辞書やシソーラスの整備が課題である。

3. デジタルアーカイブ・システムで提供するサービスに係る調査検討

3.1. デジタルアーカイブ・システムで提供するサービスに係る調査

3.1.1. 概要

本節では、国立公文書館が、平成 28 年度に全国の公文書館等を対象として実施したアンケートの結果から、それぞれの機関がデジタルアーカイブ・システムに求めるサービス・機能等について検討した。

また、合わせて、全国の公文書館等及び民間、海外の代表的なデジタルアーカイブを対象として、デジタルアーカイブ・システムの基本機能(利用者向け、業務)の動向及び特徴的事例について行った調査結果から、デジタルアーカイブ・システムで提供するサービス等について検討した。

アンケート結果からは、以下の状況が確認された。

- 目録情報の検索を支援する機能(サジェスト機能、異体字等検索機能)、動画・音声の再生機能、レファレンス等における問い合わせ機能に対して半数近い機関が必要性を示しており、利用者の利便性向上が期待されている。
- 各機能における国際標準への対応や提供データのオープンデータへの対応等、今後考慮すべきと思われる意見があった。
- 機関内部における職員向けの非公開資料の検索・閲覧機能(資料の利用制限機能)や、レファレンス業務の管理機能、利用請求記録の管理に関する機能に対するニーズが高いことが分かった。レファレンスを充実させることで利用サービス向上と利用者増を実現している事例があり、このニーズとも合致している。
- デジタルアーカイブ・システムで提供するサービス等について直接的に関与するものではないが、システムの導入・運用形態において、インフラ及びアプリケーションを共有してもよいという機関が 6 割を超えており、特にかけられる費用が 100 万円未満の機関において多く見られた。大きな予算的制約、人的制約の中で、小規模な機関ほど、低コストで運用人員を抑制できる共有型システムへの期待が大きいことがアンケートの調査結果より確認された。

デジタルアーカイブ・システムの基本機能(利用者向け、業務)の動向等調査からは、以下の状況が確認された。

- 全国の公文書館等で目録検索システムをインターネット上で公開している機関は全体の 5 割弱。
- 全国の公文書館等の半数を超える機関で画像を公開。
- 民間及び海外のデジタルアーカイブは動画・音声を含めたデジタルコンテンツを公開。
- SNS を利用した情報発信は広報ツールとしての効果が期待できる反面、不適切な情報発信への利用に留意。
- 利用者が能動的に参加する仕組みをデジタルアーカイブ・システムに実装した事例を国内外ともに確認。ボランティア等を活用して目録データの整備を進める方法の一つとして注目。

3.1.2. 全国の公文書館等へのアンケート

(1) 実施概要

デジタルアーカイブ・システムに求められるサービス・機能等の把握を目的として、国立公文書館は平成 28 年度に、全国の公文書館等及び平成 28 年度全国公文書館長会議参加機関の計 97 機関に対してアンケートを実施した。アンケートの設問は、大きく「デジタルアーカイブ・システムに求められるサービス・機能」と「望ましいシステムの運用形態」の 2 つに分けられており、また回答は、選択式及び自由記述によるものとなっている。

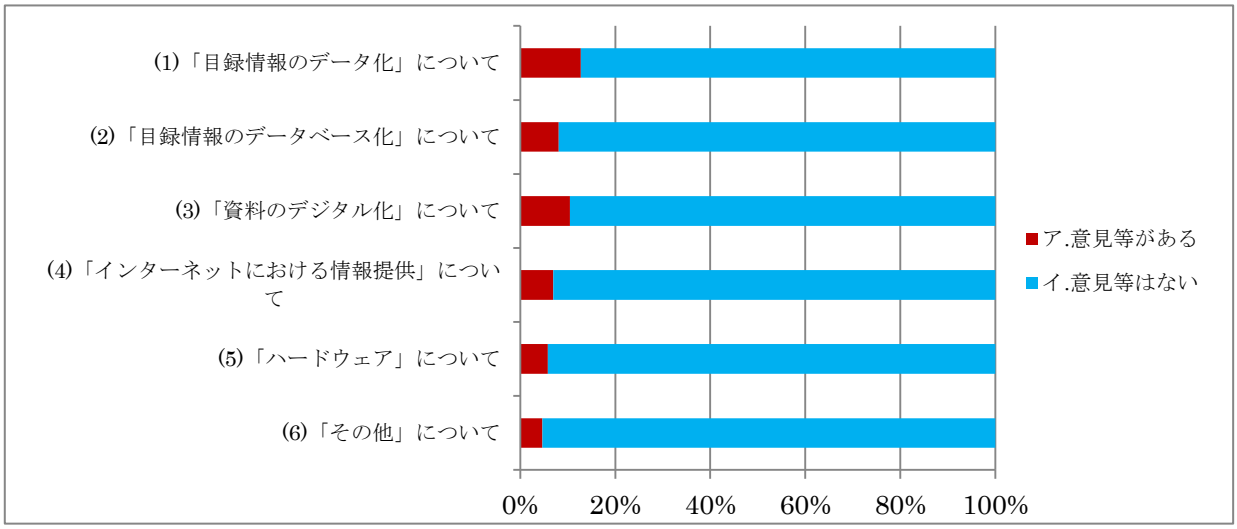
表 16 アンケートの実施状況

対象機関	97 機関(全国の公文書館等及び平成 28 年度全国公文書館長会議参加機関)
配布・回収方法	E-mail にて配布・回収
実施期間	平成 28 年 11 月
回答数	95 機関
回答率	97.9%
回答方式	選択式及び自由記述
調査内容	1. デジタルアーカイブ・システムに求められるサービス・機能 ● 「デジタルアーカイブ・システムの標準仕様書」記載の機能に関する基本的な考え方 ● 利用者がデジタルアーカイブ・システムに求めると思われる機能 ● 職員等が業務上デジタルアーカイブ・システムに求めると思われる機能 ● デジタルアーカイブ・システムを用いた利活用事例 2. 望ましいシステムの運用形態 ● 共同でデジタルアーカイブ・システムを導入・運用する場合に想定される相手先機関 ● 他機関との共有が可能と考えられるシステム構成 ● 今後デジタルアーカイブ・システムにかけることができると思われる費用 ● デジタルアーカイブ・システムの導入における制約

(2) アンケート集計結果

(ア)「デジタルアーカイブ・システムに求められるサービス・機能」の結果

設問 1 では、現在、デジタルアーカイブ・システム導入時の参考資料として、全国の公文書等に向けて国立公文書館が公表する、デジタルアーカイブ・システムの標準仕様書(以下「標準仕様書」という。)における、「目録情報のデータ化」等主な項目に対する意見の有無を確認した(図 35)。



※総回答数:86

図 35 設問 1 の回答結果

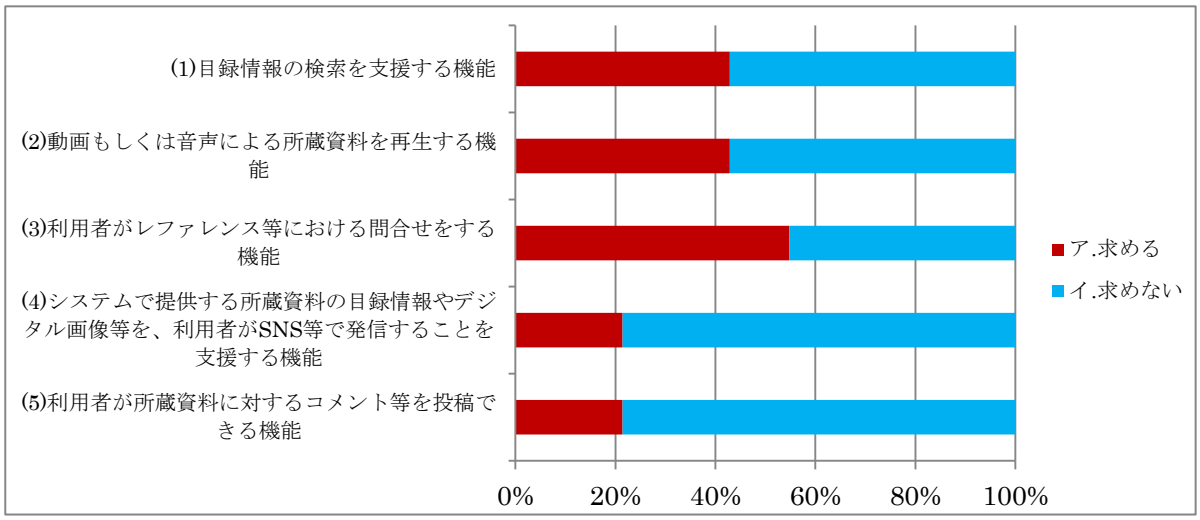
「意見がある」と回答した機関は、設問 1 の各問において 5～13%程度と少数であり、概ね標準仕様書記載の内容に対して異存はないとするものであった。しかし、少数ながらも記された意見には以下のようなものが見受けられた（表 17）。

表 17 設問 1 における主な意見

問	主な意見
「目録情報のデータ化」	ICA（国際公文書館会議）で検討されている RiC 等への対応 目録記述に係る国際標準の導入 資料群階層の構成と組織改正時における再整理の煩雑さへの対応
「目録情報のデータベース化」	オープンデータへの対応 登録情報に対する公開・非公開機能の詳細化
「資料のデジタル化」	画像ファイル形式とその推奨理由の一層の明確化 国際規格（IIIF）による画像データの相互運用への対応 画像ファイルの取り扱いに係る適切な記載
「インターネットにおける情報提供」	オープンデータへの対応 横断検索機能に係る実運用を想定した記述
「ハードウェア」	サーバ等ハードウェアの性能要件の明示 複数の機関によるシステム共有に係る考え方
「その他」	運用保守に係る仕様の明確化

また、各問を通して、標準仕様書の用語が難しいという意見があった。標準仕様書は広範な内容について専門的な用語を用いて説明しているからであると考えられるが、理解を助けるために用語集等をより平易な表現でまとめる必要があるのではないかと思われる。

設問 2 では、設問 2-1 及び設問 2-2 の二つの問を設け、利用者がデジタルアーカイブ・システムに求めると思われる機能について、そのニーズを確認した。設問 2-1 では予め用意した機能に対し、選択式による回答を求めた。



※総回答数:84

図 36 設問 2-1 の回答結果

「目録情報の検索支援機能」について、43%の機関がこの機能を「求める」と回答している。また、この項目では「求める」とした回答者に対し、求める機能を自由記述により具体的に回答することを依頼している。

上記自由記述による回答の主なものとしては、利用者に検索キーワードの候補語を提示するサジェスト機能、異体字や旧字等を考慮した検索機能について多くの機関から挙げられた。また、その他の回答として、PC 等の操作に不慣れな利用者を想定した、検索操作そのものを支援する機能について回答された機関もあった。

「動画もしくは音声による所蔵資料を再生する機能」では、上記同様に 43%の機関がこの機能を「求める」と回答しており、「利用者がレファレンス等における問合せをする機能」では 54%の機関がこの機能を「求める」と回答している。

また、「システムで提供する所蔵資料の目録情報やデジタル画像等を、利用者が SNS 等で発信することを支援する機能」と「利用者が所蔵資料に対するコメント等を投稿できる機能」はともに 22%の機関が「求める」と回答するに止まった。

「利用者が所蔵資料に対するコメント等を投稿できる機能」については、この項目では「求める」とした回答者に対し、求める機能を自由記述により具体的に回答することを依頼しており、少数ではあるが、国外の公文書館で採用されている市民アーキビスト制度のように、利用者からの提供情報による登録データの充実や利用者の拡大を期待する回答があった。その一方で、機関登録情報と利用者提供情報との区別ができるようにすべきとの意見や利用者間のトラブルが発生した場合の対応を懸念する意見があった。

また、設問 2-1 への回答において、回答機関におけるデジタルアーカイブ・システムの有無と公文書館等の区分に着目し、クロス集計を行った結果、以下の通りとなった。

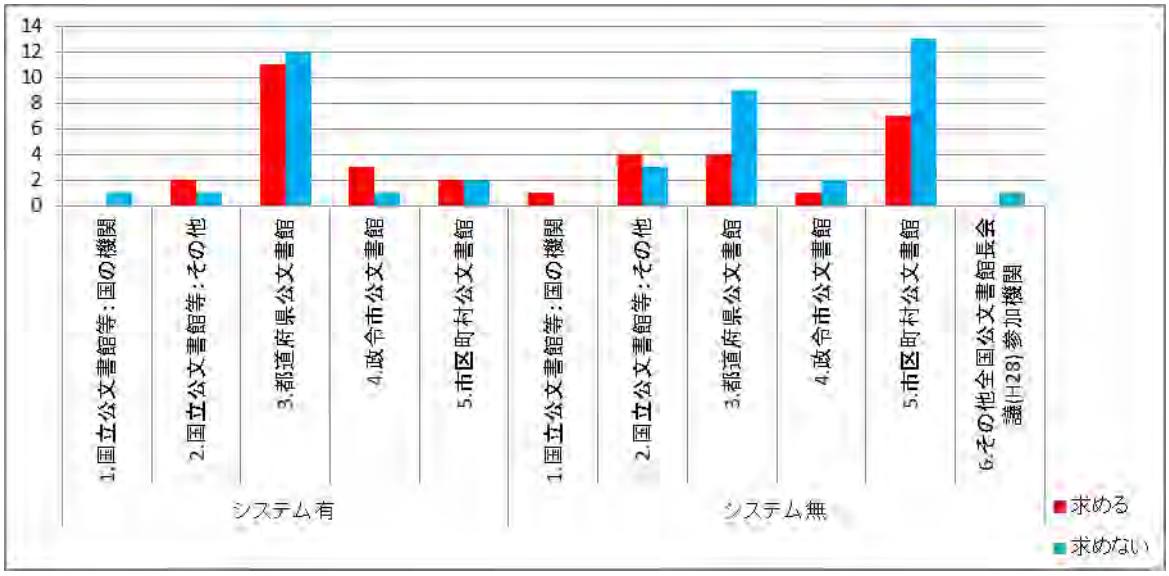


図 37 目録情報の検索を支援する機能のクロス集計結果

(1)目録情報の検索を支援する機能について、デジタルアーカイブ・システムを有さない機関においては、都道府県公文書館及び市区町村公文書館で、「求めない」を選択した機関が倍近く上回った。

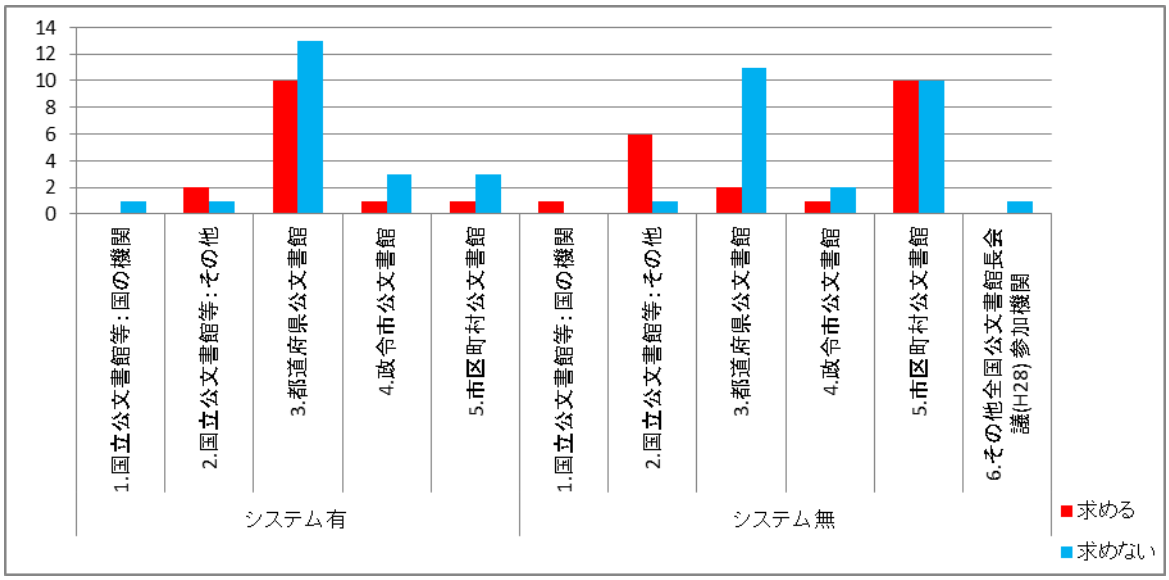


図 38 動画もしくは音声による所蔵資料を再生する機能のクロス集計結果

(2)動画もしくは音声による所蔵資料を再生する機能について、デジタルアーカイブ・システムを有する機関においては、「求めない」を選択した機関が若干上回った。しかし、同システムを有さない機関においては、都道府県公文書館で「求めない」を選択した機関が大幅に上回った。

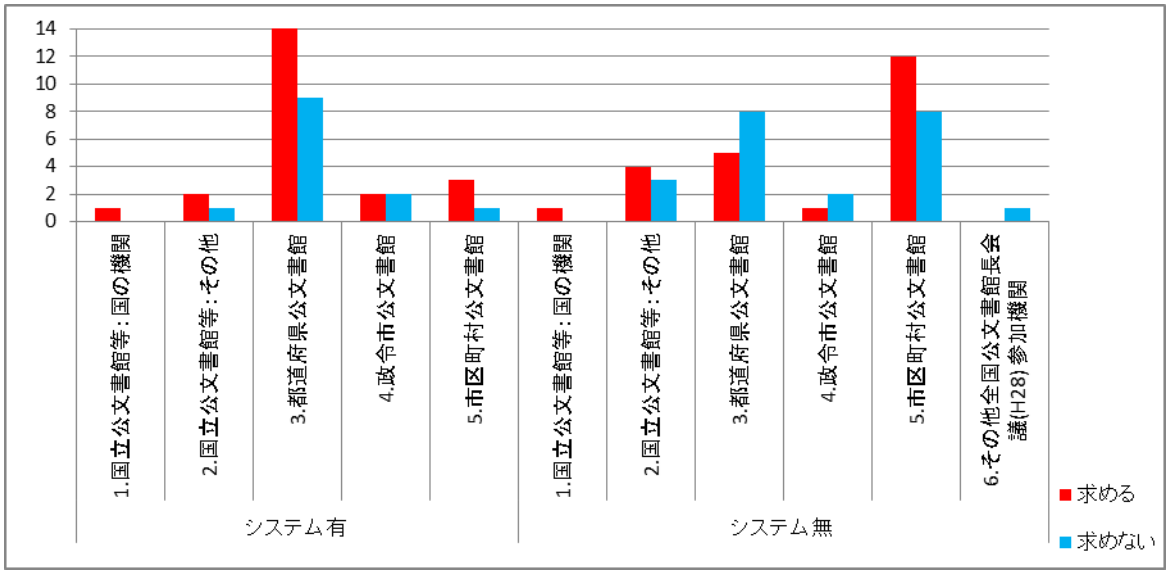


図 39 利用者がレファレンス等における問合せをする機能のクロス集計結果

(3)利用者がレファレンス等における問合せをする機能について、デジタルアーカイブ・システムの有無により、多少のばらつきがあるものの全体的に「求める」が多い傾向が見られたが、同システムを有さない都道府県公文書館では「求めない」を選択した機関が上回った。

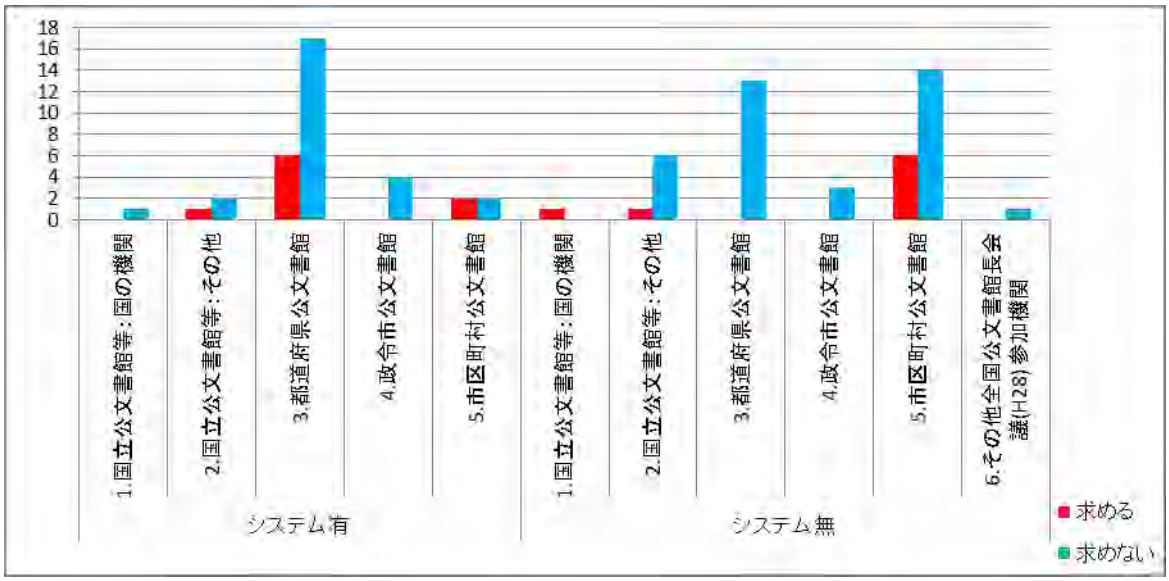


図 40 システムで提供する所蔵資料の目録情報やデジタル画像等を、利用者が SNS 等で発信することを支援する機能のクロス集計結果

(4)システムで提供する所蔵資料の目録情報やデジタル画像等を、利用者が SNS 等で発信することを支援する機能について、デジタルアーカイブ・システムの有無にかかわらず、「求めない」を選択した機関が大きく上回った。

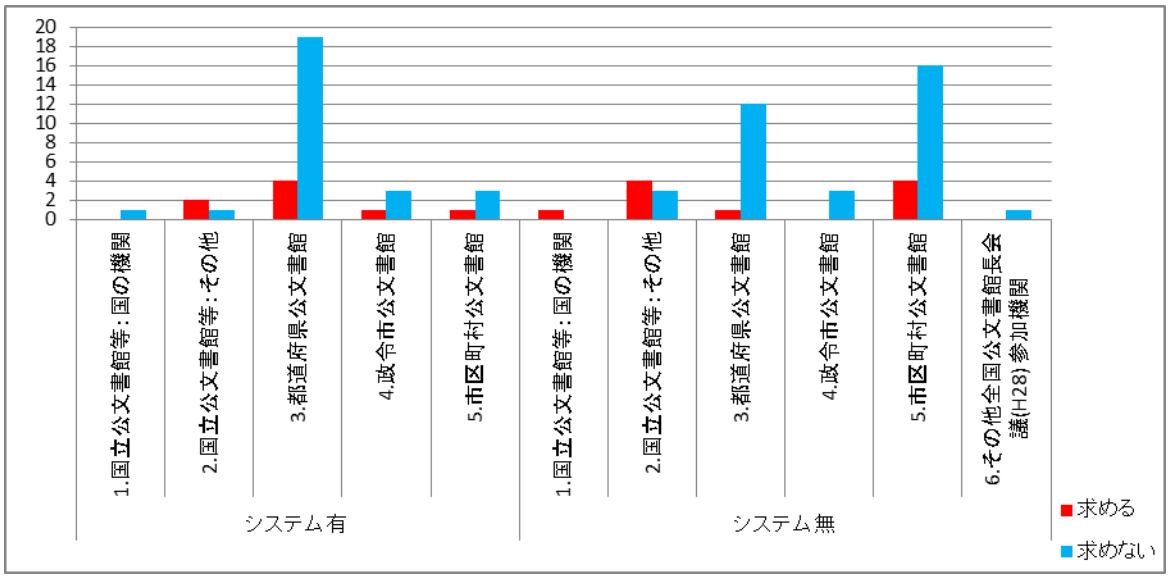
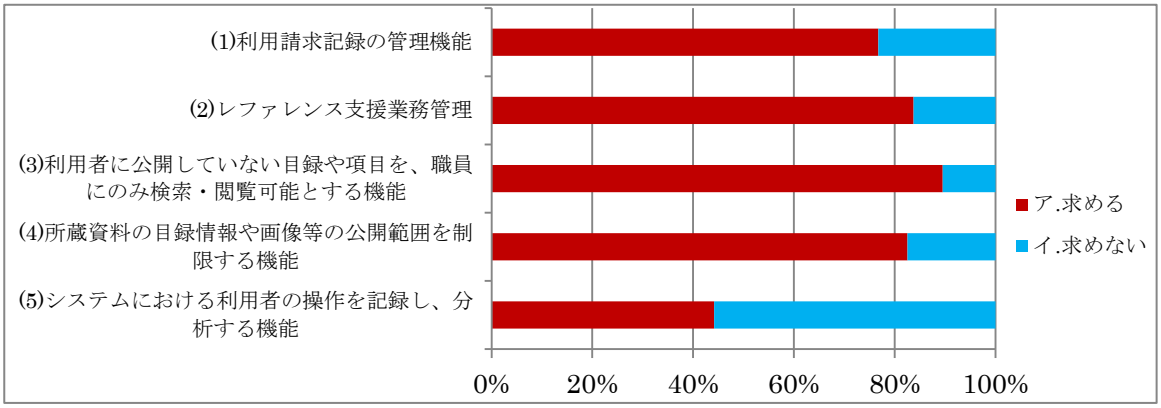


図 41 利用者が所蔵資料に対するコメント等を投稿できる機能のクロス集計結果

(5)利用者が所蔵資料に対するコメント等を投稿できる機能について、デジタルアーカイブ・システムの有無にかかわらず、「求めない」を選択した機関が大きく上回った。

設問 2-2 では、自由記述により、設問 2-1 で提示した機能以外に、利用者が求めると思われる機能について尋ねた。回答には、音声応答や操作のヘルプ機能、出典記載に対応できるようパーマリンクの提供、古文書のデジタル画像の補助機能として読み下し文の表示機能や検索結果から選択した目録情報をダウンロードする機能が挙げられた。また、引用など二次利用の制限を明示する機能として、政府標準利用規約や Creative Commons への準拠明示を求める回答があった。

設問 3 では、設問 3-1 及び設問 3-2 の二つの問を設け、職員等が業務上デジタルアーカイブ・システムに求める機能について確認した(図 42)。設問 3-1 では予め用意した機能に対し、選択式による回答を求めた。



※総回答数：86

図 42 設問 3-1 の回答結果

「利用請求記録の管理機能」については 78%の機関がこの機能を「求める」と回答しており、「レファレ

ンス支援業務管理」「利用者に公開していない目録や項目を、職員にのみ検索・閲覧可能とする機能」「所蔵資料の目録情報や画像等の公開範囲を制限する機能」については、80%を超える機関がこれらの機能を「求める」と回答している。また、「システムにおける利用者の操作を記録紙、分析する機能」は45%の機関が「求める」と回答するに止まった。

設問 3-1 について、回答機関におけるデジタルアーカイブ・システムの有無と公文書館等の区分に着目し、クロス集計を行った結果、以下の通りとなった。

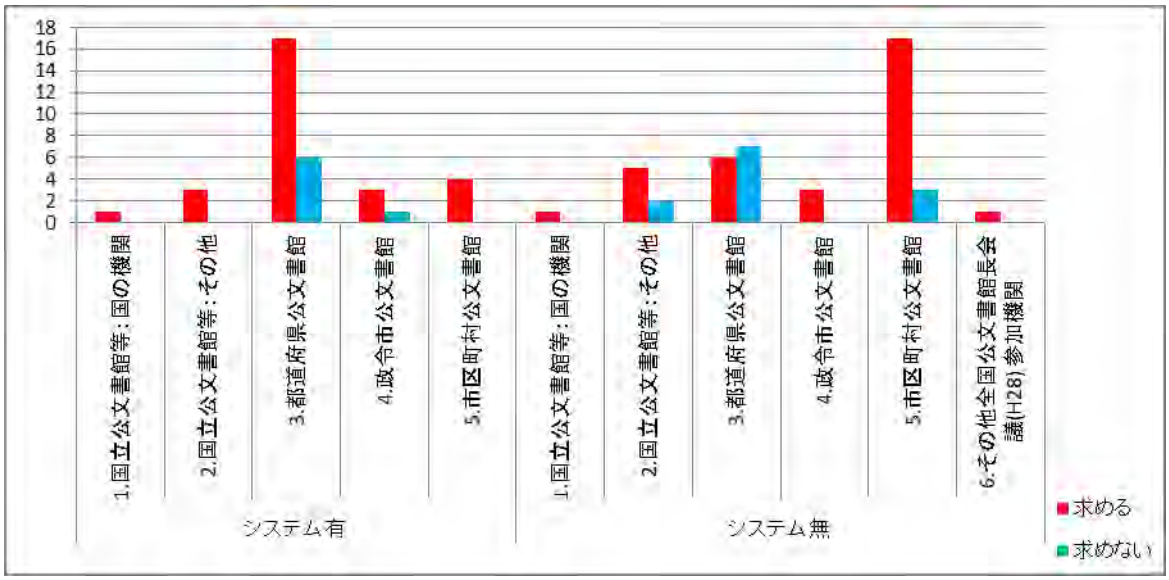


図 43 利用請求記録の管理機能のクロス集計結果

(1)利用請求記録の管理機能について、デジタルアーカイブ・システムの有無にかかわらず、「求める」と回答した機関が多かった。

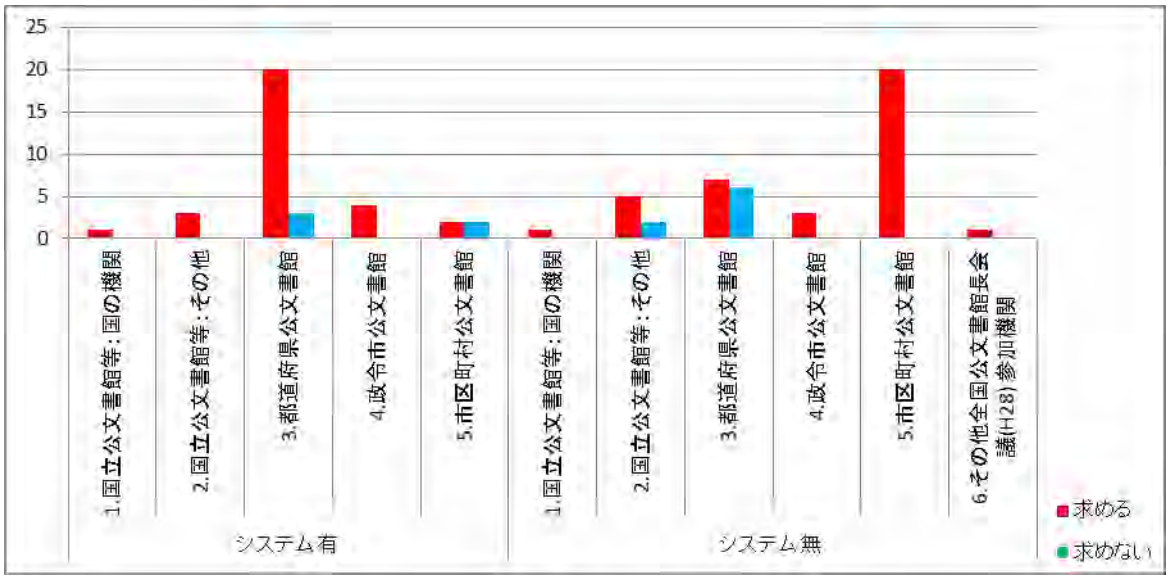


図 44 レファレンス支援業務管理のクロス集計結果

(2)レファレンス支援業務管理について、デジタルアーカイブ・システムの有無にかかわらず、「求める」と回答した機関が多かった。

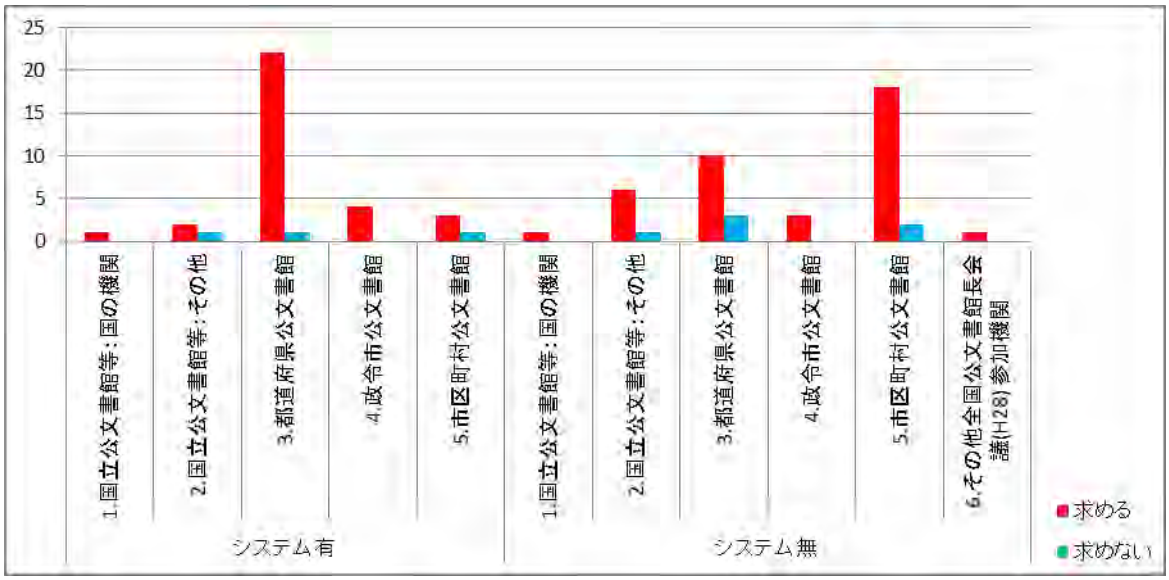


図 45 利用者に公開していない目録や項目を、職員にのみ検索・閲覧可能とする機能のクロス集計結果

(3)利用者に公開していない目録や項目を、職員にのみ検索・閲覧可能とする機能について、デジタルアーカイブ・システムの有無にかかわらず、「求める」と回答した機関が多かった。

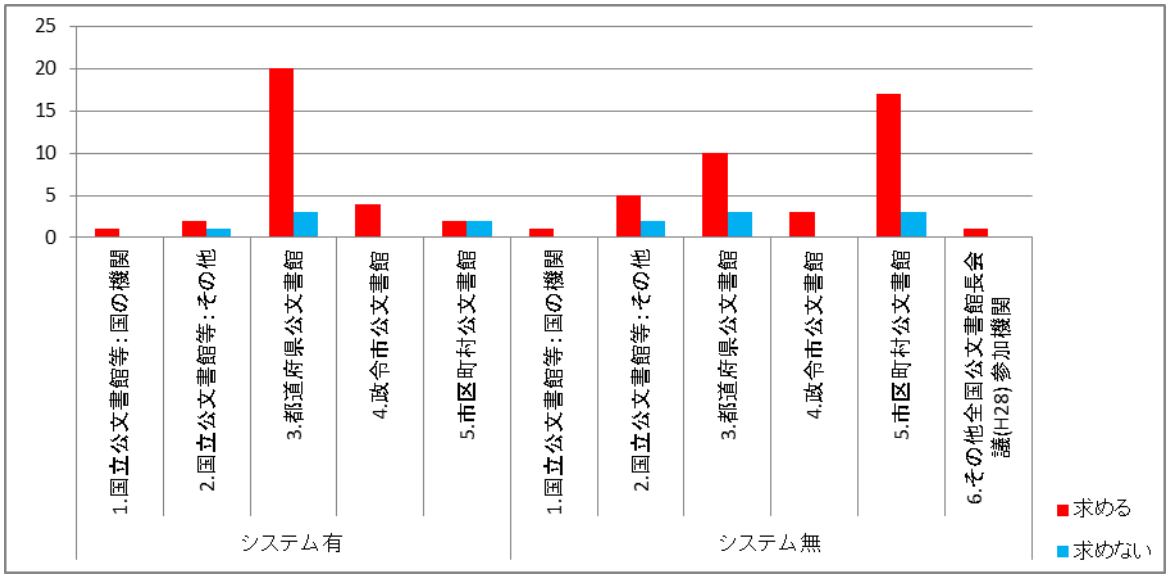


図 46 所蔵資料の目録情報や画像等の公開範囲を制限する機能のクロス集計結果

(4)所蔵資料の目録情報や画像等の公開範囲を制限する機能について、デジタルアーカイブ・システムの有無にかかわらず、「求める」と回答した機関が多かった。

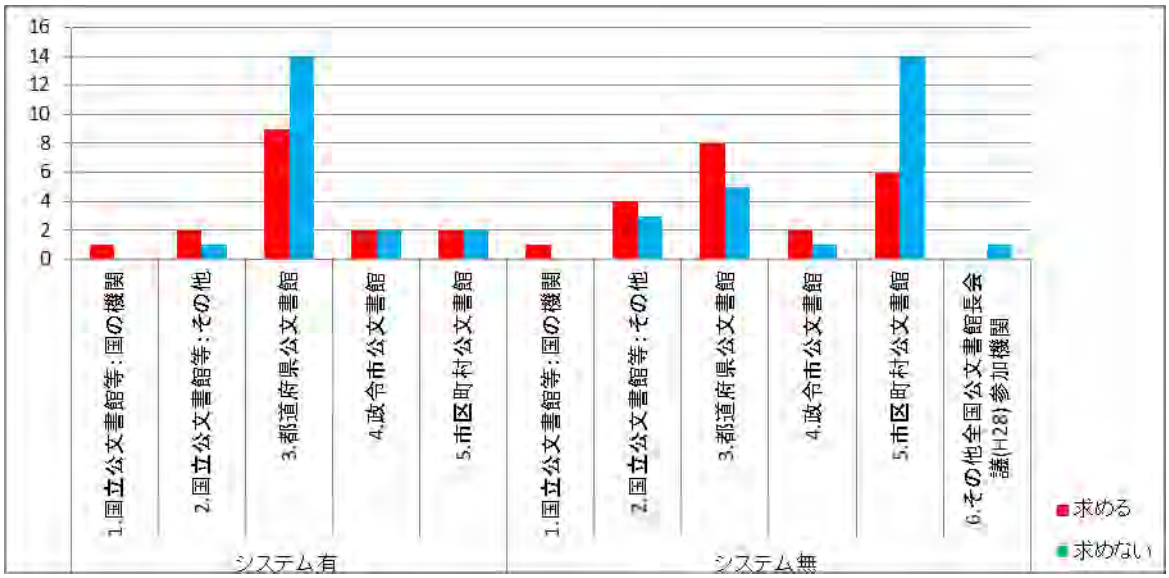


図 47 システムにおける利用者の操作を記録し、分析する機能のクロス集計結果

(5)システムにおける利用者の操作を記録し、分析する機能について、都道府県公文書館ではデジタルアーカイブ・システムを有する場合に「求めない」と回答した機関が多かったが、同システムを有する都道府県公文書館では「求める」と回答した機関が多かった。また、同システムを有さない市区町村公文書館では「求めない」と回答した機関が「求める」と回答した機関を大きく上回った。

設問 3-2 では、自由記述により、設問 3-1 で提示した機能以外に、利用者が求めると思われる機能について尋ねた。主な回答として、各機関の業務や環境に依拠する機能についてのものが挙げられたほか、「利用者に公開していない目録や項目を、職員にのみ検索・閲覧可能とする機能」について、自館だけでなく同一自治体内の他組織での利用や、寄贈者・寄贈団体など特定の利用者のみに公開したいという意見があった。

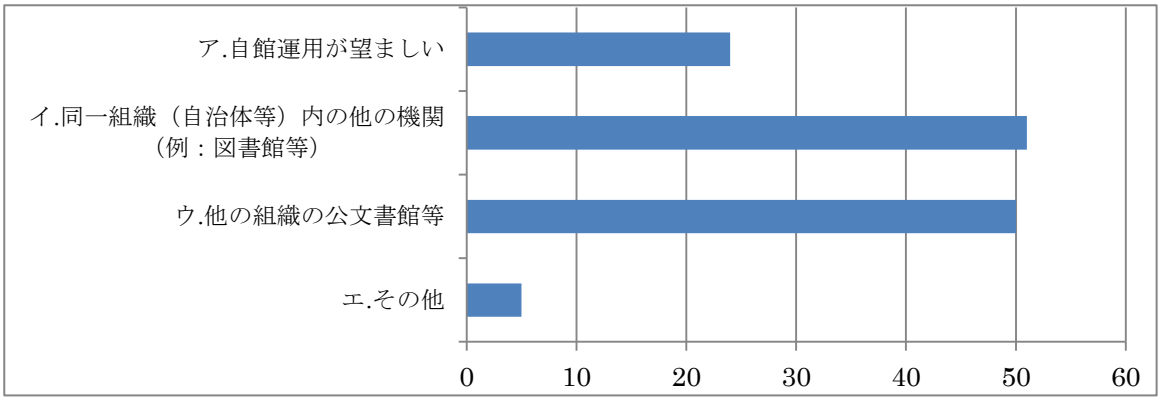
その他、システムにおける利用者の操作を記録し、分析する機能について、開示請求の対象とならない単純なアクセス件数等のログ情報で十分であるとの回答や、全ての操作を記録するようなものではなく統計的に整理できるものでよいとの回答があった。

設問 4 では、自由記述により、デジタルアーカイブ・システムを用いた利活用事例について確認した。

この設問では、目録情報の提供のほか、デジタル化資料の提供（特に原本の代替物としての提供）、メールによる利用申請に基づくデジタルデータのダウンロード提供、パーマリンク（固定 URL）を用いた各種ページでの活用といった回答等が挙げられた。

(イ) 「望ましいシステムの運用形態」の結果

設問 5 では、設問 5-1 及び設問 5-2 の二つの問を設け、他の機関と共同でデジタルアーカイブ・システムを導入・運用する場合に、想定される相手先機関について確認した。(図 48)。設問 5-1 では、「自館運用が望ましい」「同一組織（自治体等）内の他の機関」「他の組織の公文書館等」「その他」の四つからなる選択肢により回答を求めた。



※総回答数：84

図 48 設問 5 の回答結果

回答は複数選択可としたことから、その合計数は全回答機関数を上回るものとなっているが、「同一組織(自治体等)内の他の機関」(39%)と「他の組織の公文書館等」(38%)を想定する機関が多いこと、また、「自館運用が望ましい」(18%)とする機関も少なからずあることが分かった。

設問 5-1 について、回答機関におけるデジタルアーカイブ・システムの有無と公文書館等の区分に着目し、クロス集計を行った結果、以下の通りとなった。

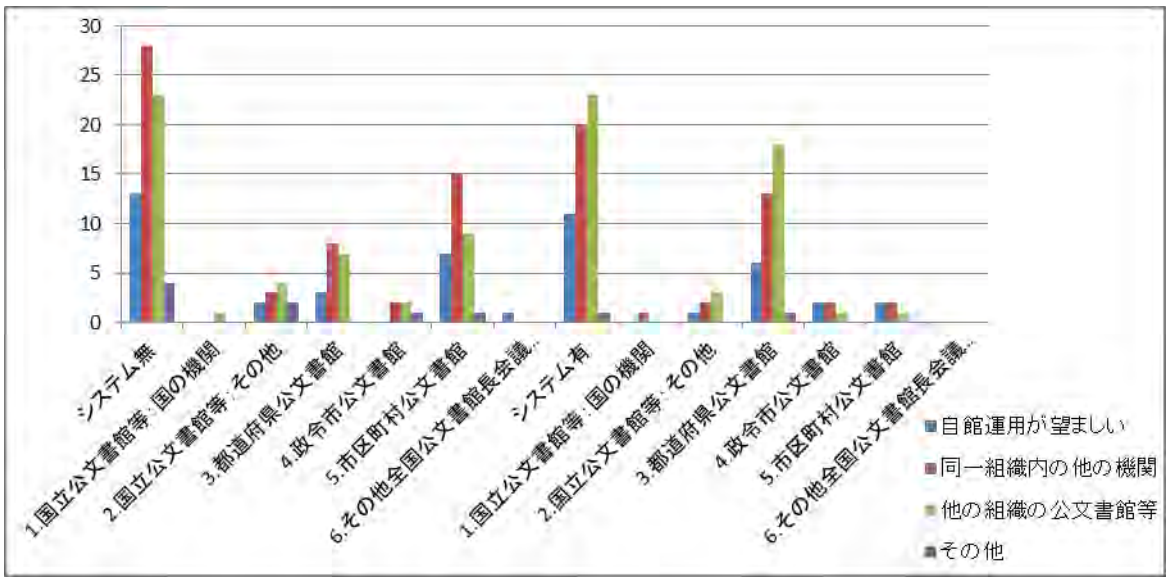


図 49 他館との共同利用についてのクロス集計結果

同システムを有する機関では、同一組織内の他の機関との共同利用を選択した機関が多く、同システムを有さない機関では他の組織の公文書館等を選択した機関が多かった。一方で、公文書館等の区分ごとに見ると都道府県公文書館では同システムを有する機関では他の組織の公文書館等を選択した機関が上回り、市区町村公文書館では同システムを有さない機関で同一組織内の他の機関を選択した機関が上回った。

設問 5-2 では、自由記述により、設問 5-1 で選択した回答に対する理由について尋ねた。

「自館運用が望ましい」と回答した機関は、システムを導入済みの機関が多く、現行システムの機能やデータ移行の制約、柔軟に運用できること、データを外部に持ち出せない等が理由として挙げられた。

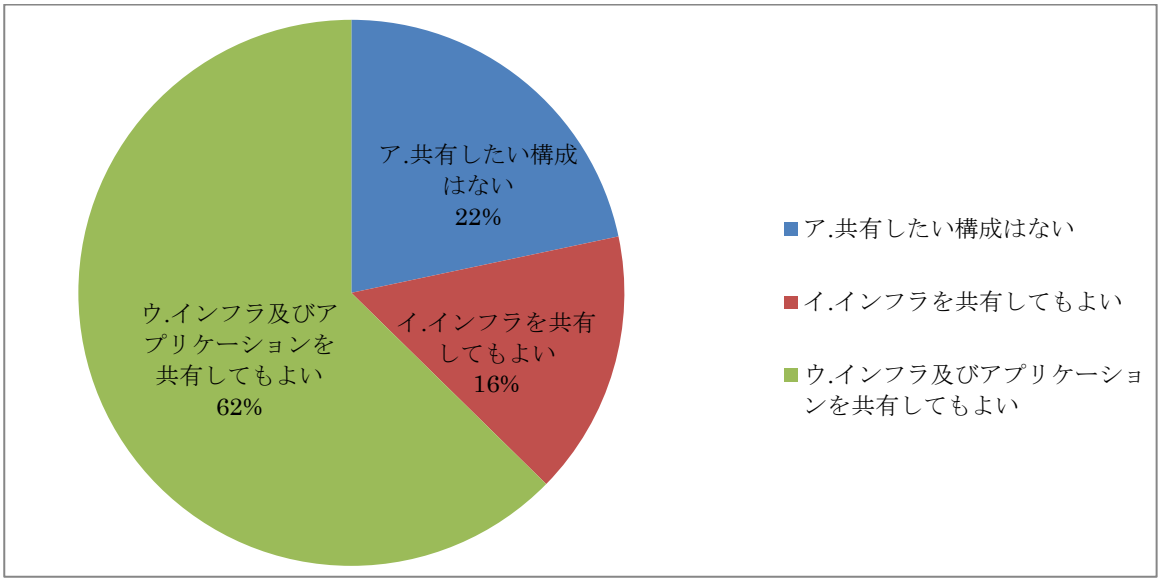
「同一組織(自治体等)内の他の機関」と回答した機関では、すでに自治体内で構築済みのシステムを利用している場合や、将来的に図書館・博物館といった機関との共通システムを利用する計画であるという回答があった一方で、図書館とのデータ構造や検索方法の違いを指摘する意見を添えた機関もあった。

「他の組織の公文書館等」と回答した機関では、同じ公文書館であることでデータ構造や機能が似通っており連携がしやすい、コストの軽減、利用者の利便性向上という内容が理由として挙げられている。

また、「その他」と回答した機関からは、各機関の実情に照らし合わせたシステムの在り方に触れた内容の回答が挙げられた。

設問 6 では、設問 6-1 及び設問 6-2 の二つの問を設け、デジタルアーカイブ・システムの導入・運用において、他の機関との共有が可能と考えられるシステム構成について確認した。(図 50)。

設問 6-1 では、「共有したい構成はない」「インフラを共有してもよい」「インフラ及びアプリケーションを共有してもよい」の三つからなる選択肢により回答を求めた。



※総回答数:83

図 50 設問 6 の回答結果

「共有したい構成はない」と回答した機関が 21%、「インフラを共有してもよい」と回答した機関が 16%とそれほど多くないのに対し、「インフラ及びアプリケーションを共有してもよい」と回答した機関が 6 割近くに上った。

設問 6-1 について、回答機関におけるデジタルアーカイブ・システムの有無と公文書館等の区分に着目し、クロス集計を行った結果、以下の通りとなった。

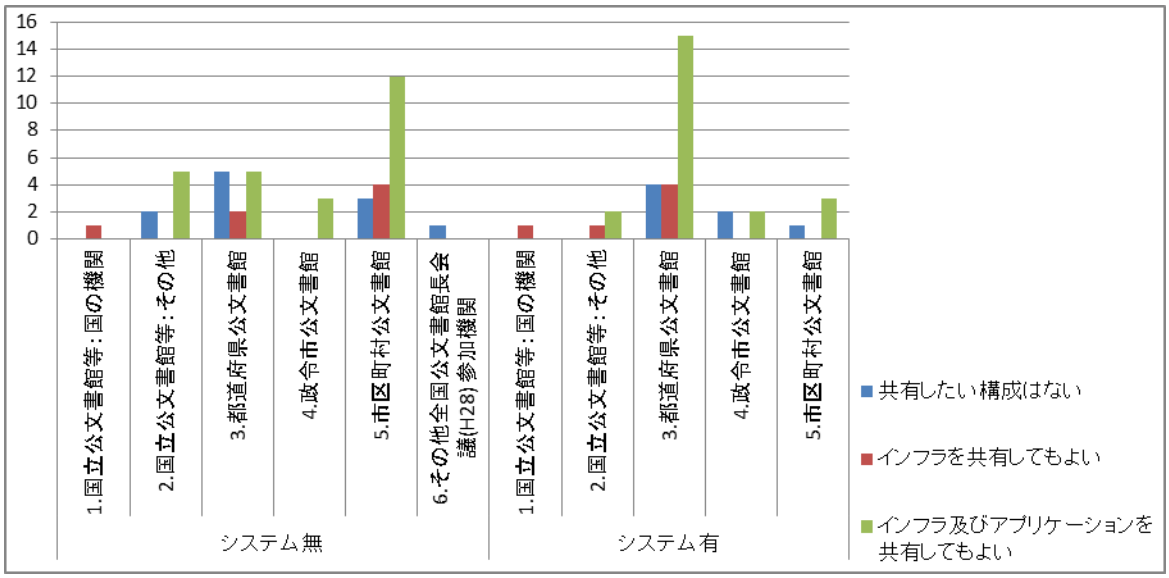


図 51 共有可能と考えられるシステム構成についてのクロス集計結果(システム有無・公文書館等の区分)

公文書館等の区分とデジタルアーカイブ・システムの有無を合わせて集計すると、同システムを有さない市区町村公文書館と同システムを有する都道府県公文書館で、「インフラ及びアプリケーションを共有してもよい」と回答した機関が多かった。また、設問 7 の(1)システム構築にかけられる費用の回答と合わせて集計すると、100 万円未満と回答した機関で、「インフラ及びアプリケーションを共有してもよい」と回答した機関が突出して多かった。

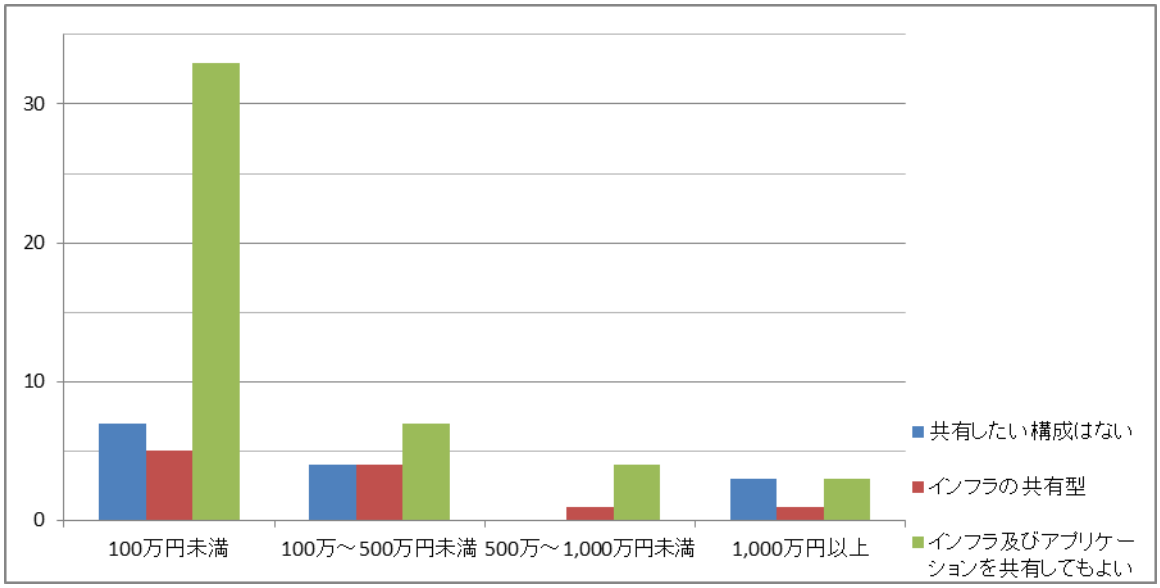


図 52 共有可能と考えられるシステム構成についてのクロス集計結果(システム構築ににかけられる費用)

設問 6-2 では、自由記述により、設問 6-1 で選択した回答に対する理由について尋ねた。

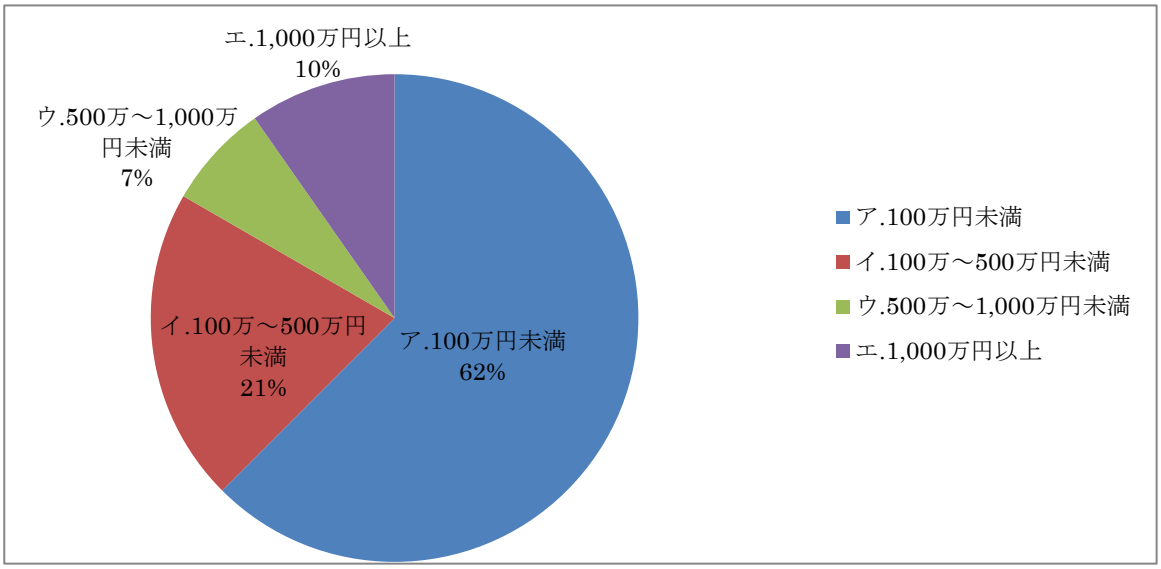
「インフラ及びアプリケーションを共有してもよい」理由として、予算確保が難しい中で構築・運用コストを軽減できることを挙げる意見が最も多かった。

「インフラを共有してもよい」という機関では、インフラのコスト軽減を図りながらもアプリケーションには独自性が必要と考える意見が見られた。「共有したい構成はない」という機関では、主にセキュリティ上の

制約や機能・データ項目の柔軟性が失われることへの懸念が挙げられた。また、共有することに対して、関係部署・機関との調整や費用分担を課題とする意見があった。

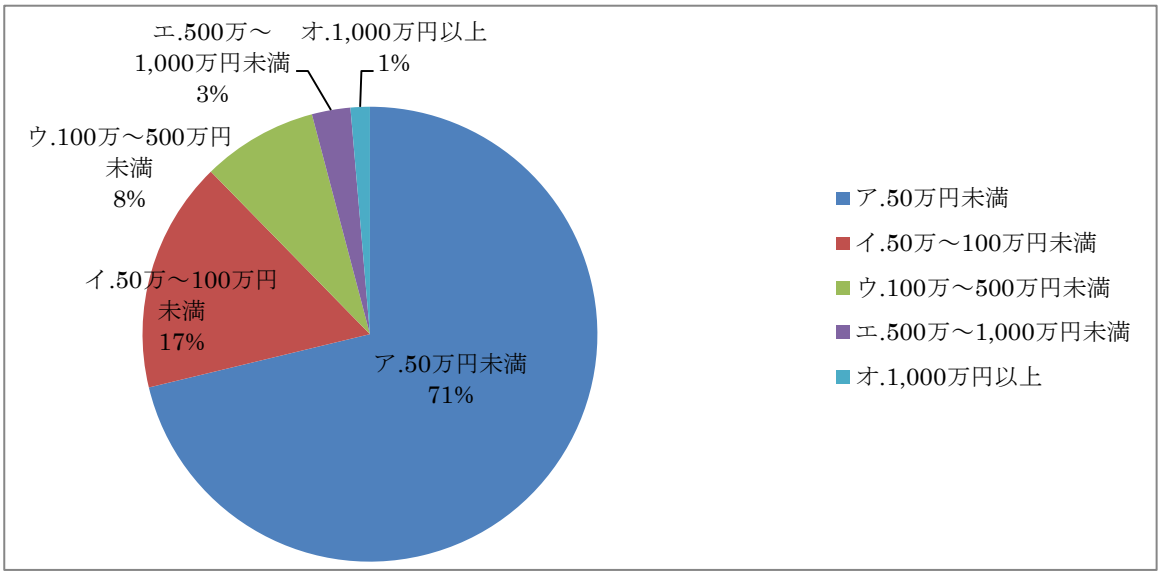
設問 6-1 と設問 6-2 の回答を踏まえると、予算が少なく、小規模な機関ほど、安価で導入でき、システム運用の人員を抑制できることが期待される「インフラ・アプリケーション共有型システム」を望んでいると考えられる。

設問 7 では、今後デジタルアーカイブ・システムにかけることができると思われる費用について、システム導入時に発生する費用(図 53)と保守・運用等費用(図 54)に分けて確認した。



※総回答数:72

図 53 システム導入時にかけられる費用の回答結果



※総回答数:73

図 54 システム保守・運用等にかけられる費用の回答結果

システム導入時にかけられる費用では、100 万円未満と回答した機関が 6 割に上り、保守・運用等費

用では、50 万円未満と回答した機関が 7 割に上った。

設問 7 について、回答機関におけるデジタルアーカイブ・システムの有無と公文書館等の区分に着目し、クロス集計を行った結果、以下の通りとなった。

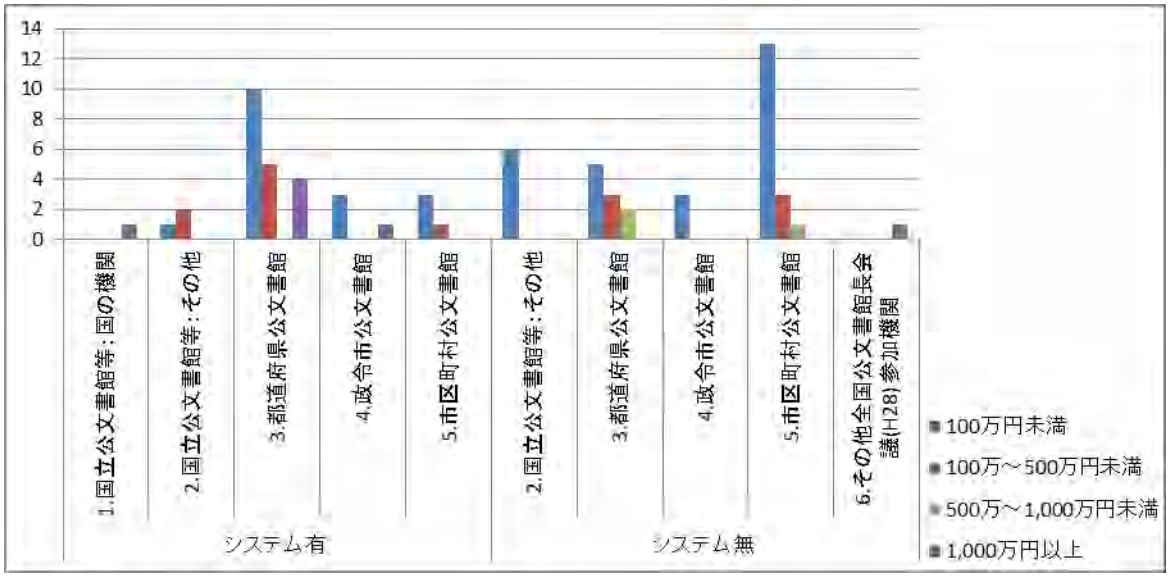


図 55 構築時にかけられると思われる額についてのクロス集計結果

(1)デジタルアーカイブ・システム導入時にかけられると思われる費用については、現時点における同システムの有無にかかわらず、100 万円未満が圧倒的に多かった。全体として予算確保の難しさが顕著に表れているが、公文書館等の区分別では市区町村公文書館において 100 万円未満という回答の割合が最も高く、機関規模が小さくなるほど、予算確保がさらに困難であるものと考えられる。

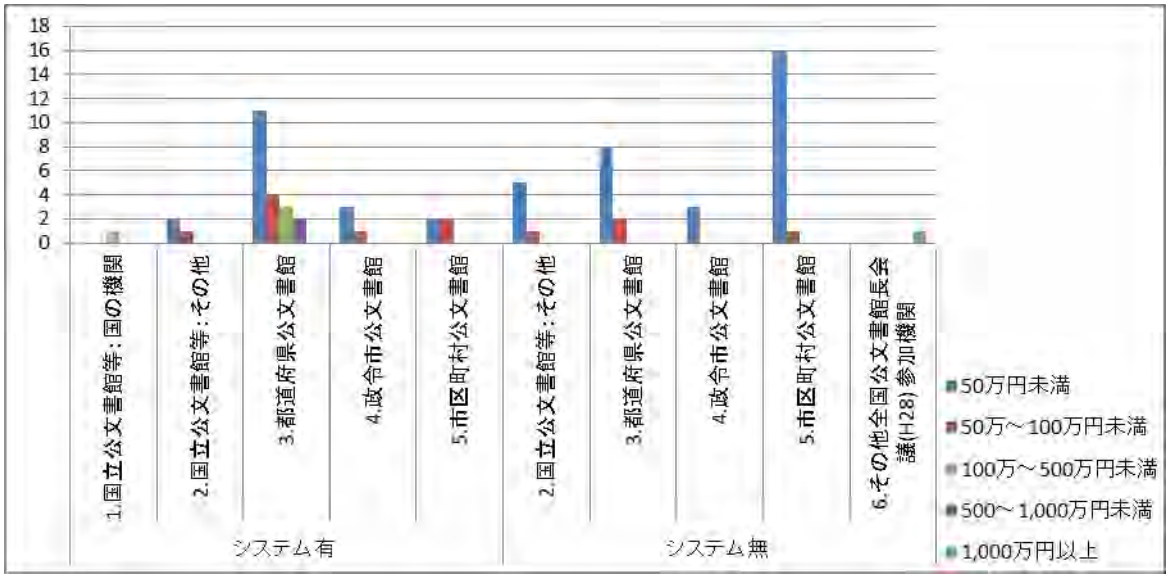


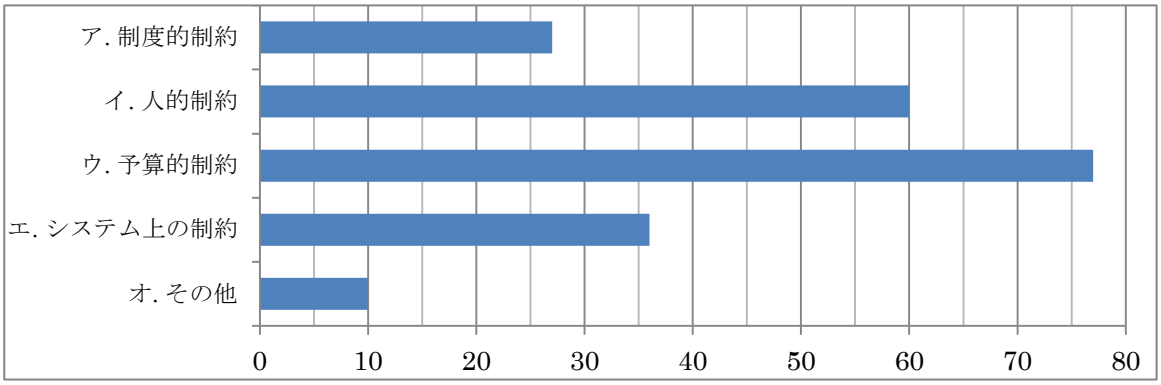
図 56 保守・運用等費用(年額)にかけられると思われる額についてのクロス集計結果

(2)デジタルアーカイブ・システムの保守・運用等にかけられると思われる費用についても、現時点における同システムの有無にかかわらず、50 万円未満が圧倒的に多かった。システム導入時と同様に、公文書館等の区分別では市区町村公文書館において 50 万円未満という回答の割合が最も高く、機関規模が

小さくなるほど、予算確保がシステム導入時よりもさらに困難であるものと考えられる。

設問 8 では、設問 8-1 及び設問 8-2 の二つの問を設け、デジタルアーカイブ・システムの導入における制約について確認した。

設問 8-1 では、「制度的制約」「人的制約」「予算的制約」「システム上の制約」「その他」の五つからなる選択肢により回答を求めた。(図 57)



※総回答数:83

図 57 設問 8 の回答結果

回答は複数選択可としたことから、その合計数は全回答機関数を上回るものとなっているが、最も多かったのは「予算的制約」とした機関であり、次が「人的制約」とした機関であった。また、「制度的制約」「システム上の制約」と回答した機関も少なからずあることが分かった。

設問 8-1 について、回答機関におけるデジタルアーカイブ・システムの有無と公文書館等の区分に着目し、クロス集計を行った結果、以下の通りとなった。

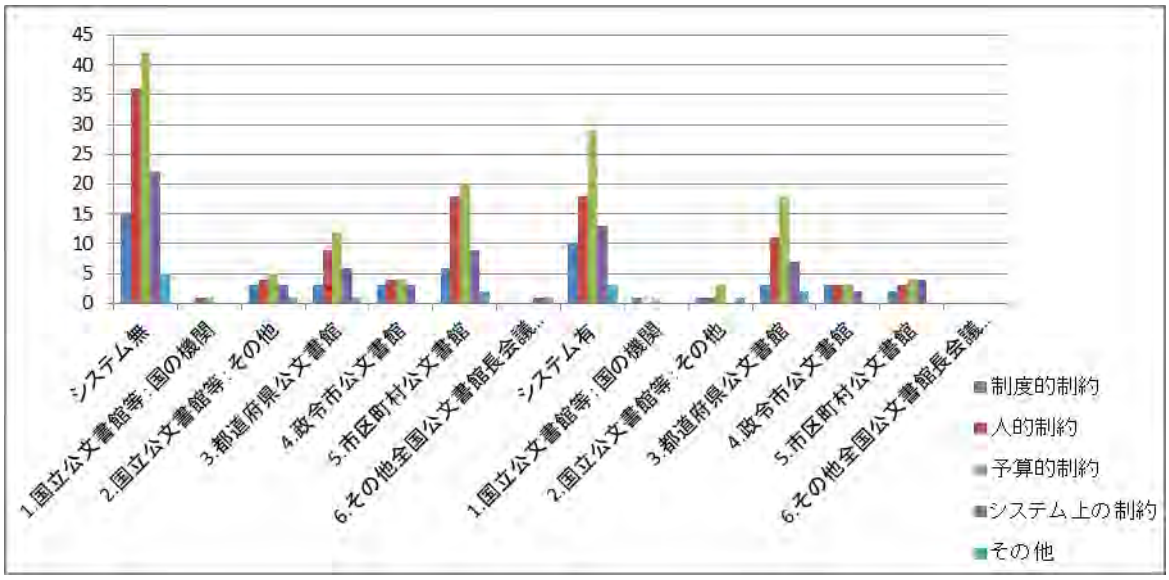


図 58 導入における制約についてのクロス集計結果

現時点における同システムの有無にかかわらず、「予算的制約」が最も多く、次が「人的制約」であった。市区町村公文書館の場合は同システムを有さない機関が多いため、「予算的制約」と「人的制約」の 2 点

を制約として挙げる機関が多く見られる。システム化にあたっての予算確保と少人数による運用の実現が大きな課題であるといえる。

設問 8-2 では、自由記述により、設問 8-1 で選択した制約について具体的な記述を求めた。

「予算的制約」では、予算獲得にあたり費用対効果の説明が求められるという記述が多く、「人的制約」では専門的知識を有する職員を配置することが困難であるという記述が多く見受けられた。「システム上の制約」では自治体内のサーバを使うことによる制限や現行システムとの整合性、「制度的制約」では、自治体のセキュリティポリシーを挙げる記述が多く見受けられた。

また、「その他」と回答した機関からは、各機関独自の実情に基づく記述が見受けられた。

設問 8-1 と設問 8-2 を踏まえると、デジタルアーカイブ・システムの導入時と保守・運用時の両者に係る中長期的な支援が求められているものと考えられる。

(3) 考察

(ア) 設問ごとの考察

- 設問 1 について

設問 1 は現在、国立公文書館が公表している標準仕様書の主な項目における記載内容への意見を確認するものである。本設問における各項目に対して意見を記した機関は全体から見て少数だった。しかし、示された意見の中には各機能における国際標準への対応や提供データのオープンデータへの対応等、今後考慮すべきと思われるものも確認された。

また、そのほかに、標準仕様書の用語が難しいという意見があった。標準仕様書は広範な内容について専門的な用語を用いて説明しているからであると考えられるが、理解を助けるために用語集等をより平易な表現でまとめる必要があるのではないかと思われる。

- 設問 2-1 について

(1)目録情報の検索を支援する機能

回答機関のうち、半数近い機関がこの機能の必要性を示した。また、具体的な機能として、「検索キーワード入力時におけるサジェスト機能」や「異体字、旧字体等を考慮した検索機能」等が挙げられた。

(2)動画もしくは音声による所蔵資料を再生する機能

回答機関のうち、半数近い機関がこの機能の必要性を示した。もはや所蔵資料として動画や音声を扱うことは特別なことではなく、これらの資料を利用者へ提供する機能が当然のように求められているものと思われる。

(3)利用者がレファレンス等における問合せをする機能

回答機関のうち、半数を超える機関がこの機能の必要性を示した。多くの機関がレファレンス等の利用者対応を重要な業務として位置づけるとともに、システム化することによる利用者に対する利便性の向上を期待しているものと思われる。

(4)システムで提供する所蔵資料の目録情報やデジタル画像等を、利用者が SNS 等で発信することを支援する機能

この機能の必要性を示した機関は回答機関の 22%程度に止まった。多くはないが、この機能を

有用なものとして捉えている機関があることを示しているものと思われる。

(5)利用者が所蔵資料に対するコメント等を投稿できる機能

この機能の必要性を示した機関は回答機関の 22%程度に止まった。また、多くはないが、この機能により、利用者からの提供情報による登録データの充実等を期待する回答が寄せられた。

- 設問 2-2 について

音声応答や操作のヘルプ機能、出典記載に対応できるようパーマリンクの提供、古文書のデジタル画像の補助機能として読み下し文の表示機能や検索結果から選択した目録情報をダウンロードする機能が挙げられた。また、引用など二次利用の制限を明示する機能として、政府標準利用規約や Creative Commons への準拠明示を求める回答があった。

- 設問 3-1 について

(1)利用請求記録の管理機能

全体の 78%と回答機関の多くがこの機能の必要性を示した。所蔵資料に対する利用請求業務を重要な業務として位置づけるとともに、システム化することによる業務改善を少なからず求めているものと思われる。

(2)レファレンス支援業務管理

全体の 80%を超える回答機関がこの機能の必要性を示した。多くの機関がレファレンス等の利用者対応を重要な業務として位置づけるとともに、システム化することによる業務改善を少なからず求めているものと思われる。

(3)利用者に公開していない目録や項目を、職員にのみ検索・閲覧可能とする機能

全体の 80%を超える回答機関がこの機能の必要性を示した。多くの機関において、利用者に公開する対象ではない情報を持っていることが示されているものと考えられる。また、それらの情報を特定の職員が検索・閲覧できることが業務上必要であるものと思われる。

(4)所蔵資料の目録情報や画像等の公開範囲を制限する機能

全体の 80%を超える回答機関がこの機能の必要性を示した。多くの機関において、所蔵資料に関する情報のうち、公開を制限している情報を持っていることが示されているものと考えられる。

(5)システムにおける利用者の操作を記録し、分析する機能

回答機関のうち、半数近い機関がこの機能の必要性を示した。また、都道府県公文書館ではデジタルアーカイブ・システムを有する場合に「求めない」と回答した機関が多かったが、同システムを有さない都道府県公文書館では「求める」と回答した機関が多かった。さらに同システムを有さない市区町村公文書館では「求めない」と回答した機関が「求める」と回答した機関を大きく上回った。

- 設問 3-2 について

各機関の業務や環境に依拠する機能についてのものが挙げられたほか、「利用者に公開していない目録や項目を、職員にのみ検索・閲覧可能とする機能」について、自館だけでなく同一自治体内の他組織での利用や、寄贈者・寄贈団体など特定の利用者のみに公開したいという意見があった。その他、紙資料からのテキスト起こしをインターネット上でボランティアを募り、協働作業として実施できる機能への要望があった。データの品質や責任を明確にしたうえで、近年広まりつつあるクラウドソーシングを活用することも低コストでデータ化を行う選択肢としては有効と考えられる。

- 設問 4 について

目録情報の提供のほか、デジタル化資料の提供（特に原本の代替物としての提供）、メールによる利用申請に基づくデジタルデータのダウンロード提供、パーマリンク（固定 URL）を用いた各種ページでの活用といった回答等が挙げられた。

- 設問 5 について

他の機関と共同でデジタルアーカイブ・システムを導入・運用する場合の想定される相手先機関について、同システムを有する機関では、「同一組織（自治体等）内の他の機関」との共同利用を望む回答が多く、同システムを有さない機関では「他の組織の公文書館等」を望む回答が多かった。

- 設問 6 について

全体として、「インフラ及びアプリケーションを共有してもよい」という回答が多かった。特に「システム構築にかけられる費用」に対して、100 万円未満と回答した機関でもインフラ及びアプリケーションを共有してもよいという回答が突出していた。

予算が少なく、小規模な機関ほど、安価で導入でき、システム運用の人員を抑制できることが期待される「インフラ・アプリケーション共有型システム」を望んでいると考えられる。

- 設問 7 について

デジタルアーカイブ・システム導入時にかけられると思われる費用については、現時点における同システムの有無にかかわらず 100 万円未満が圧倒的に多かった。全体として予算確保の難しさが顕著に表れているが、公文書館等の区分別では市区町村公文書館において 100 万円未満という回答の割合が最も高く、機関規模が小さくなるほど予算確保がさらに困難であるものと考えられる。

デジタルアーカイブ・システムの保守・運用等にかけられると思われる費用についても、現時点における同システムの有無にかかわらず 50 万円未満が圧倒的に多かった。システム導入時と同様に、公文書館等の区分別では市区町村公文書館において 50 万円未満という回答の割合が最も高く、機関規模が小さくなるほど、予算確保がシステム導入時よりもさらに困難であるものと考えられる。

- 設問 8 について

現時点におけるデジタルアーカイブ・システムの有無にかかわらず、「予算的制約」が最も多く、次が「人的制約」であった。

市区町村公文書館の場合は同システムを有さない機関が多いため、上記の二点を制約として挙げる機関が多く見られる。設問 7 での考察と同様に、システム化にあたっての予算確保と少人数による運用の実現が大きな課題であるといえる。

また、設問 8-1 と設問 8-2 を踏まえると、デジタルアーカイブ・システムの導入時と保守・運用時の両者に係る中長期的な支援が求められているものと考えられる。

（イ）全体の考察

本アンケートの結果を通して、第一に全国の公文書館等、とりわけ市区町村レベルではシステム化の予算と人員の確保が大きな課題となっていると考えられる。また、その課題の解決策としてアプリケーションも含めた安価な共同利用型システムが期待されていると思われる。インフラとアプリケーションを共有してもよいと回答した機関が 6 割以上である一方、単館運用を希望する機関も一定数あり、機関の要望に応じて適切なシステム構成や運用形態を選択できることが必要である。

利用者向けの機能としては、サジェスト機能や異体字等検索機能など検索支援機能の整備が期待されている。利用者の利便性向上のため、想定する利用者に合わせた検索支援機能を充実させることが重要である。また、動画や音声の提供機能へのニーズが多い。もはやマルチメディア資料を扱うことが特別なことではなく、これらの資料を利用者へ提供する機能が当然のように求められていると考えられる。

一方で業務面では、利用請求記録の管理機能のほか、レファレンス業務を管理する機能や、業務上必要となる非公開データの検索・閲覧に係る機能（資料の利用制限機能）が求められている。特に共同利用型システムにおける要件を検討する際には、これらが重要な要素であると考えられる。

3.1.3. デジタルアーカイブの基本機能(利用者向け)

現状、全国の公文書館等がインターネットで提供している利用者向けのサービスは、デジタルアーカイブ・システムで必要なサービスであると考えられる。本項では、まず全国の公文書館等が提供している基本機能(目録検索、画像閲覧、利用請求等)の基礎調査を実施し、さらに民間や海外のデジタルアーカイブの事例を調査し、デジタルアーカイブにおいて求められる基本機能・サービスについて検討する。

(1) 全国の公文書館等の事例

デジタルアーカイブ・システムで提供するサービスの在り方を確認するため、全国の公文書館等における、デジタルアーカイブ・システムやホームページなど、インターネットを通じた情報提供の状況を調査した。また、目録検索機能、画像の配信・閲覧機能、横断検索といったデジタルアーカイブ・システムにおける利用者向けの基本機能の事例について調査を行った。

調査概要

調査目的	インターネットによる情報提供の状況把握、デジタルアーカイブの基本機能(利用者向け)の事例調査
調査対象	国立公文書館のホームページにおいて、国内外や類縁機関等、国立公文書館に関係のあるウェブサイトを集めたリンク集[104]に示されている、国内109 機関
調査方法	上記機関のホームページ、目録検索システム等にアクセス(平成 28 年 12 月 19 日現在)
調査内容	・目録の公開状況 ・目録検索機能の状況 ・画像の公開状況 ・利用請求の支援状況 ・スマートデバイスへの対応状況 ・その他特徴的な事例

(ア) 目録の公開状況

調査結果を表 18 に示す。「検索システム有」の場合は PDF、Excel 等で目録を公開していても「PDF、Excel 等で公開」には含めていない。109 機関のうち、47%にあたる 51 機関が目録検索システムにより目録を公開していた。PDF、Excel 等により目録情報を公開している機関を含めると、71%が目録をインターネット上で公開している。インターネット上で目録を公開している機関は、都道府県公文書館では 89%であるが、市区町の公文書館では 42%であり、自治体規模により公開状況に大きな隔たりがある。

---

104 http://www.archives.go.jp/links/ (2017 年 1 月 24 日アクセス)

表 18 目録の公開状況

公文書館等の区分	検索システム有	PDF、Excel等で公開	なし	**総計**
1.国立公文書館に類する機能を有するものとして、公文書管理法に基づき定められた施設	6	6	2	**14**
2.国の保存利用機関	2	0	1	**3**
3.類縁機関・大学アーカイブズ等	9	2	9	**20**
4.都道府県公文書館	24	9	4	**37**
5.政令市	5	3	1	**9**
6.市区町	5	6	15	**26**
総計	**51**	**26**	**32**	**109**

（イ）目録検索機能の状況

調査結果を表 19、表 20 に示す。目録検索システムにより目録を公開している 51 機関を対象として、目録検索機能が以下の機能を備えているか調査した。

- 「キーワード検索」機能 ･･･ キーワードを指定して検索する機能
- 「詳細検索」機能 ･･･ キーワードの他に条件を指定して検索する機能
- 「階層検索」機能 ･･･ 公文書等に関する階層構成を選択しながら目録を検索する機能

キーワード検索機能は 40 機関（78%）に備わっており、詳細検索機能は 39 機関（76%）である。キーワード検索と詳細検索の導入割合が同程度であるのは、目録検索の利用者が、キーワードでなく項目を指定して特定の資料を探すような使い方へのニーズが高いためと推測される。

表 19 キーワード検索機能の導入状況

公文書館等の区分	キーワード検索有	なし	**総計**
1.国立公文書館に類する機能を有するものとして、公文書管理法に基づき定められた施設	5	1	**6**
2.国の保存利用機関	1	1	**2**
3.類縁機関・大学アーカイブズ等	7	2	**9**
4.都道府県公文書館	20	4	**24**
5.政令市	3	2	**5**
6.市区町	4	1	**5**
総計	**40**	**11**	**51**

表 20 詳細検索機能の導入状況

公文書館等の区分	詳細検索有	なし	総計
1.国立公文書館に類する機能を有するものとして、公文書管理法に基づき定められた施設	5	1	6
2.国の保存利用機関	2	0	2
3.類縁機関・大学アーカイブズ等	8	1	9
4.都道府県公文書館	18	6	24
5.政令市	3	2	5
6.市区町	3	2	5
総計	**39**	**12**	**51**

各機関のデジタルアーカイブ・システムでは、キーワード検索や項目指定検索といった基本的な機能は網羅されている。検索項目として指定できる項目はタイトル(簿冊名、件名、資料名等)、作成年・年代・年度、識別番号(文書番号、簿冊番号、請求記号等)が多い。

階層検索については検索システムを有する 45 機関[105]中 23 機関で採用されているが、ツリー階層型のユーザーインターフェースを有しているのはそのうち 12 機関である。ツリー階層型のユーザーインターフェースを備える事例として京都大学大学文書館の画面例を図 59 に示す。画面左側に階層ツリーが表示されており、「+」をクリックすると下位の階層が展開され、直観的に操作できる。

図 59 階層検索の事例(京都大学大学文書館)

階層は公文書等の場合は移管元機関・部局により作成されているが、アンケートでの意見でも挙げら

105 類縁機関・大学アーカイブズ等で多岐に渡るデータベースを運用している機関は対象外とした。

れた通り、組織改正が行われた場合に階層の付け替えや移行を行わなければならない可能性があり、運用上で容易に階層構造の変更や組織変遷を管理できる仕組みが必要と考えられる。また、古文書の場合は所蔵者を、刊行物の場合は分類や形態を階層として扱っている事例が多く、階層構造は扱う資料の種別に応じて柔軟に設定できることが必要と考えられる。

辞書(同義語やシソーラス)による検索機能を備えているのは 6 機関であり、目録検索システムを有する機関でも 12%のみである。

検索支援手段として、辞書を活用している事例(アジア歴史資料センターの五十音検索[106])、別途整備されたコンテンツから目録や画像に誘導する事例(国立公文書館の太政類典の構成[107]、省庁組織変遷図[108]、京都府立京都学・歴彩館(旧京都府立総合資料館)の「京の百年」[109])がある。いずれも画面からマウスクリック操作のみで目録詳細や画像へリンクする仕組みとなっており、利用者のキーワード入力を必要としないユーザーインターフェースを備えている点が特徴である。

(ウ) 画像の公開状況、画像の配信・閲覧機能

調査結果を表 21 に示す。対象は 109 機関とし、目録検索システムから直接連携しないケース(画像データのコレクションが目録とは別に用意されている等)、資料紹介や展示紹介等の形で一部資料の画像が閲覧できる場合も画像公開に含めることとした。

画像を公開している公文書館等は 61 機関(56%)であり、公文書館等の区分ごとでは市区町の公文書館が 27%、その他の公文書館等では 56%～71%となっており、自治体規模による差が見られる。11 機関では JavaScript や FLASH を用いた高精細画像ビューアが提供されている。36 機関では JPEG 形式や PDF 形式の画像をブラウザ上でそのまま表示する仕組みとなっている。

表 21 画像の公開状況

公文書館等の区分	画像公開	なし	**総計**
1.国立公文書館に類する機能を有するものとして、公文書管理法に基づき定められた施設	10	4	**14**
2.国の保存利用機関	2	1	**3**
3.類縁機関・大学アーカイブズ等	14	6	**20**
4.都道府県公文書館	23	14	**37**
5.政令市	5	4	**9**
6.市区町	7	19	**26**
総計	**61**	**48**	**109**

高精細画像ビューアの例として、秋田県立公文書館では、絵図等の高精細画像をマウスの操作のみ

106 https://www.jacar.archives.go.jp/aj/meta/syllabary (2017 年 1 月 24 日アクセス)
107 https://www.digital.archives.go.jp/dajou/ (2017 年 1 月 24 日アクセス)
108 https://www.digital.archives.go.jp/hensen/ (2017 年 1 月 24 日アクセス)
109 http://www.archives.kyoto.jp/websearchpe/hundred/ (2017 年 1 月 24 日アクセス)

でスムーズに拡大・縮小したり、移動したりして閲覧することができる。

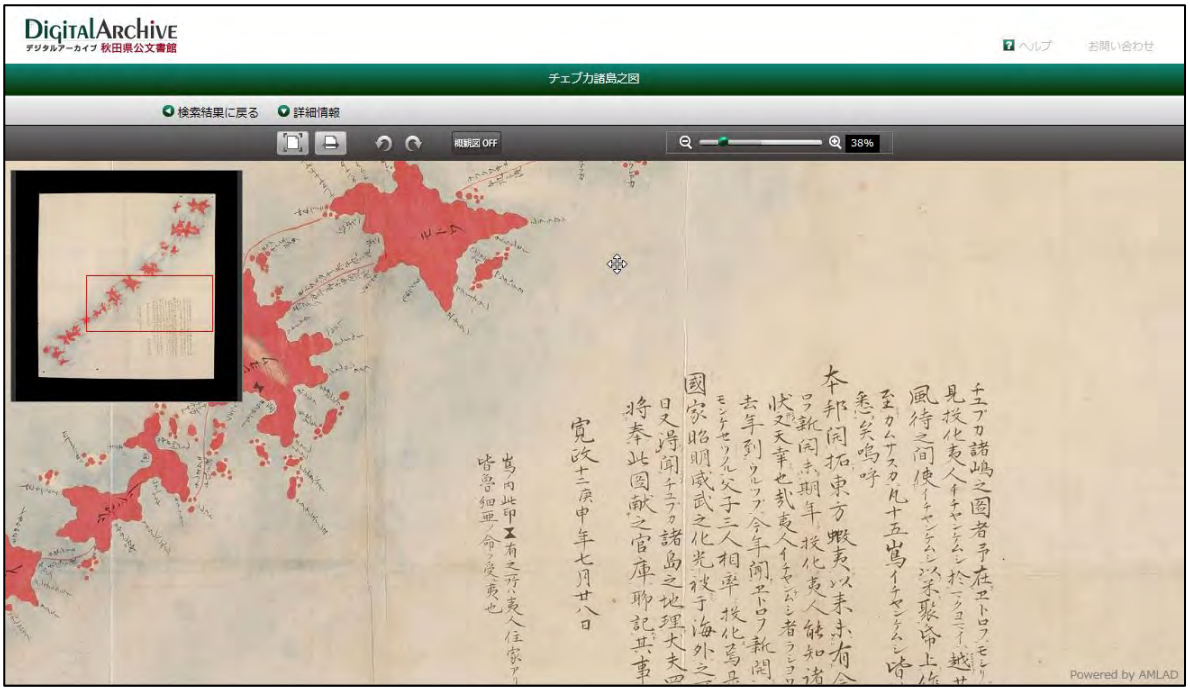


図 60 公文書館等における画像ビューアの事例(秋田県立公文書館)

(エ) 横断検索機能

横断検索では、全国の文書館等を対象とした事例(国立公文書館)、自治体内の類縁機関を対象とした事例(秋田県立公文書館)が見られる。行政文書(公文書等)と古文書、図書といった種別ごとにデータベースが分かれている場合に、全体を横断的に検索できる仕組みを備えている事例もある(群馬県立文書館、茨城県立歴史館等)。アンケート結果では、特に自治体の公文書館等で、資料が自治体内の複数の組織に分散されている場合に横断検索を行えることで、利用者の利便性向上につながるという意見があった。

国立公文書館の横断検索では、国立公文書館/アジア歴史資料センターを含めた公文書館等 11 機関と、国立国会図書館、国立情報学研究所を合わせた 13 機関のデータベースが対象となっている。

(オ) その他の機能

利用請求の支援については、109 機関のうち 53 機関で、デジタルアーカイブ・システムの他、ホームページから利用請求書のダウンロードが可能となっており、利用者による利用請求の支援機能を提供している。これらのサービスの多くは、PDF や Word 等の形式で利用請求書の様式をダウンロードし、利用者が必要事項を記載する方式が取られているが、目録検索の結果から請求書に必要な情報を自動的に出力する仕組みを備えている例(埼玉県立文書館等)や、自治体の電子申請システムから利用請求を行える機関もある(鳥取県立公文書館等)。

目録検索システムを持つ 51 機関のうち、目録検索システムがスマートフォンに最適化された表示を確認できたのは 7 機関であり、スマートフォンによる利用者への対応はあまり進んでいない状況である。

(カ) 特徴的な事例

基礎調査の中で特徴的であった事例を以下に挙げる。

- Facebook の公式ページや Twitter の公式アカウントで、イベントの告知や開館時間の案内といった広報情報や、新たに公開した資料や時事に関連する資料を紹介している機関(東京都公文書館、尼崎市立地域研究史料館等)があった。
- マスコットキャラクター「もんじょくん」と対話形式で資料を検索できたり、子どもたちにもわかりやすく文書館(アーカイブズ)の仕事を説明する「ジュニアページ」を用意したりしている機関(藤沢市文書館)があった。

(キ) 考察

デジタルアーカイブ・システムで利用者が資料を探す検索手段は目録データであり、キーワード検索や詳細検索といった基本的な検索機能は最低限必要な機能である。階層検索も資料の構成を辿りながら検索を行えるという点で利用者の利便性を高める機能であるが、資料の特性に応じて階層情報を柔軟に持てるようにする必要がある。

利便性向上という点では辞書(同義語、シソーラス)による検索機能、アンケートの結果からキーワードのサジェスト機能が必要であると考えられる。また、キーワードや階層による検索以外に、テーマや分類といった様々な切り口で資料へ辿り着ける検索支援手段のコンテンツを充実させることが有益である。

横断検索は、地理的に近接する公文書館等の機関間で所蔵する資料をまとめて検索することができ、利用者の利便性向上に資する機能である。離れた機関であっても、1 回の検索で公文書館等の資料を検索できることは利便性向上を図る重要な手段となる。

各機関のデジタルアーカイブでも画像配信・閲覧機能の整備は進んでいるが、公開されている画像データの規模、閲覧機能のレベルは機関により異なっている。

動画、音声をデジタルアーカイブ上で配信している例は現時点ではわずかであるが、デジタル化された資料の受入が今後進むことで増加すると予想される。

(2) 民間の事例

デジタルアーカイブ・システムで提供する利用者向け基本機能の在り方を確認するため、民間で提供されているデジタルアーカイブについて、利用者向けの機能やサービスの状況を調査した。

調査対象のうち、J-DAC は人文社会科学系貴重史資料のオンラインプラットフォームとして、丸善雄松堂株式会社と大日本印刷が共同開発し 2012 年に立ち上げたデジタルアーカイブであり、全国各地の大学、図書館等で大切に保存されてきた貴重資料をデジタルアーカイブ化し配信することによって、研究資源の共有化と円滑な利活用を実現しつつ、貴重資料の保全を推進することを目的としたサービスである。

また、ADEAC は、TRC-ADEAC 株式会社が制作・運営する、デジタルアーカイブの検索・閲覧を行うためのクラウド型プラットフォームシステムであり、自治体や公共図書館、大学図書館等各機関が所蔵する資料をデジタル化し、インターネットを通して閲覧可能としたサービスを提供している。また、デジタル化・検索等の方式として、2010(平成 22)年～2012(平成 24)年度における東京大学史料編纂所社会連携研究部門「図書館等所蔵資料の調査・整備研究」の成果を採用している。

調査概要

調査目的	民間のデジタルアーカイブで提供される利用者向けの機能やサービスの状況を把握する
調査対象	民間のデジタルアーカイブとして、以下を対象とした。 ・J-DAC (Japan Digital Archive Center)[110] ・ ADEAC (A system of Digitalization and Exhibition for Archive Collections)[111]
調査方法	・対象のホームページ等にアクセスし、機能等を調査する
調査内容	・検索機能、画像配信・閲覧機能、横断検索機能の状況 ・その他特徴的な事例

（ア）目録検索機能の状況

- J-DAC

「日本政治史」・「日本経済史・経営史」・「日本近代文学」・「古典籍」の 4 種類のコレクションに複数のデータベースを用意している。7 つのデータベースでキーワード検索が行えるが、メタデータはコレクション・データベースにより異なるため、その特性に合わせた検索項目や検索方法が用意されている。たとえば企業資料統合データベースでは企業名や業種、フリーワード等による検索が行えるが、太宰治自筆資料集データベースでは初出テキストからのキーワード検索のみである。

---

110 丸善雄松堂が運営するデジタルアーカイブ。
http://j-dac.jp/top/default/index.html （2017 年 1 月 24 日アクセス）

111 TRC-ADEAC 株式会社が制作・運営する、デジタルアーカイブの検索・閲覧を行うためのクラウド型プラットフォームシステム。2010 年度から 2012 年度に行われた東京大学史料編纂所社会連携研究部門の成果をコンテンツとして公開している。
https://trc-adeac.trc.co.jp/ （2017 年 1 月 24 日アクセス）

図 61 民間のデジタルアーカイブ事例(J-DAC 企業史料統合データベース検索画面)

図 62 民間のデジタルアーカイブ事例(J-DAC 太宰治自筆資料集 検索画面)

- ADEAC

全国の図書館等の機関の所蔵資料を「歴史資料」・「古典籍」・「行政資料」・「個人資料」・「新聞・広報」・「美術資料」というコレクションに分類している。メタデータは「識別番号」・「史資料番号」・「管理記号」・「所蔵機関(コレクション)」・「史料群(フォンド)」・「分類名(シリーズ)」・「副史料群(サブフォンド)」・「文書(アイテム・【タイトル】)」・「タイトル読み」などであり、階層構造を意識した記述内容となっている[112]。

48 のアーカイブが公開されており、アーカイブが複数のデータベースに分かれているものもある。多くのデータベースでキーワード検索が行えるようになっているが、データベースにより検索できる対象と内容は異なっている。

112 『2010 年度-2012 年度東京大学史料編纂所社会連携研究部門 研究成果報告書』2013 年 3 月

図 63 民間のデジタルアーカイブ事例(ADEAC 函館市 函館市地域史料アーカイブのキーワード検索画面)

図 64 民間のデジタルアーカイブ事例(ADEAC 西尾市岩瀬文庫 古典籍書誌データベースのキーワード検索画面)

（イ）画像の配信・閲覧機能

- J-DAC

J-DAC では古典資料から近現代の企業関連資料まで扱っている資料が多岐に渡っているが、画像は PDF 形式で配信されているものが多い。一部の資料では JPEG 形式で専用ビューアにより閲覧できるよ

うになっており、資料の目次を左側に表示し、閲覧するページへナビゲートする機能が特徴である。

- ADEAC

ADEAC も多様な資料を扱っており、資料の特色に応じた特徴のあるビューアが提供されている。たとえば、古文書原本と翻刻文を重ねて表示するビューアや、土器などを 3D 画像として閲覧できるビューアが用意されている。

（ウ）横断検索機能

- ADEAC

前述の ADEAC では、登録されているアーカイブ・データベースを対象とした横断検索が行えるようになっている。対象となるアーカイブ・データベースは地域・都道府県単位でまとめられており、階層的なユーザーインターフェースにより選択することができる。

図 65 民間のデジタルアーカイブ事例(ADEAC 横断検索画面)

（エ）考察

民間のデジタルアーカイブでは、取り扱う資料が多岐に渡っており、検索やデジタルコンテンツの配信・閲覧方法もその特性に応じたものが提供されている。画像ビューアでは複数資料（古文書原本と翻刻文）を重ねて表示するなど特徴的な事例があり、公文書館等において提供する機能・サービスの参考となるものがあるが、コストや技術的な課題等、実現可能性について今後検討が必要と考えられる。

(3)　海外の事例

デジタルアーカイブ・システムで提供する利用者向け基本機能の在り方を確認するため、海外で提供されているデジタルアーカイブについて、利用者向けの機能やサービスの状況を調査した。

調査概要

調査目的	海外のデジタルアーカイブで提供される利用者向けの機能やサービスの状況を把握する
調査対象	海外のデジタルアーカイブとして、以下を対象とした。 ・米国国立公文書記録管理局(National Archives and Records Administration, NARA)[113] : National Archives Catalog ・英国国立公文書館(The National Archives United Kingdom, TNA)[114] : Discovery ・オーストラリア国立公文書館(The National Archives of Australia, NAA)[115] : RecordSearch ・Europeana[116] : Europeana Collections
調査方法	・対象機関のホームページ等にアクセスし、機能等を調査する
調査内容	・検索機能、画像配信・閲覧機能、横断検索機能の状況 ・その他特徴的な事例

(ア) 目録検索機能の状況

- NARA

NARA が提供する National Archives Catalog の検索機能として、"Basic Search"と"Advanced Search"が用意されている。"Basic Search"はキーワードの入力欄が 1 つのみの単純な検索画面である。"Advanced Search"はタイトル(Title)、キーワード(Search Term)、記録グループ番号/コレクション ID(Record Group Number / Collection ID)、人名または組織名(Person or Organization Name)、日付(Date)、資料種別(Type of Archival Material)等の項目を指定した検索が行える(図 66)。

113 https://www.archives.gov/ (2017 年 1 月 24 日アクセス)
114 http://www.nationalarchives.gov.uk/ (2017 年 1 月 24 日アクセス)
115 http://www.naa.gov.au/ (2017 年 1 月 24 日アクセス)
116 Europeana は欧州連合の欧州委員会が公開している、デジタルアーカイブ群のアグリゲータであり、デジタル化された文化遺産を統合的に検索することができるポータルサイトである。
http://www.europeana.eu/portal/en (2017 年 1 月 24 日アクセス)

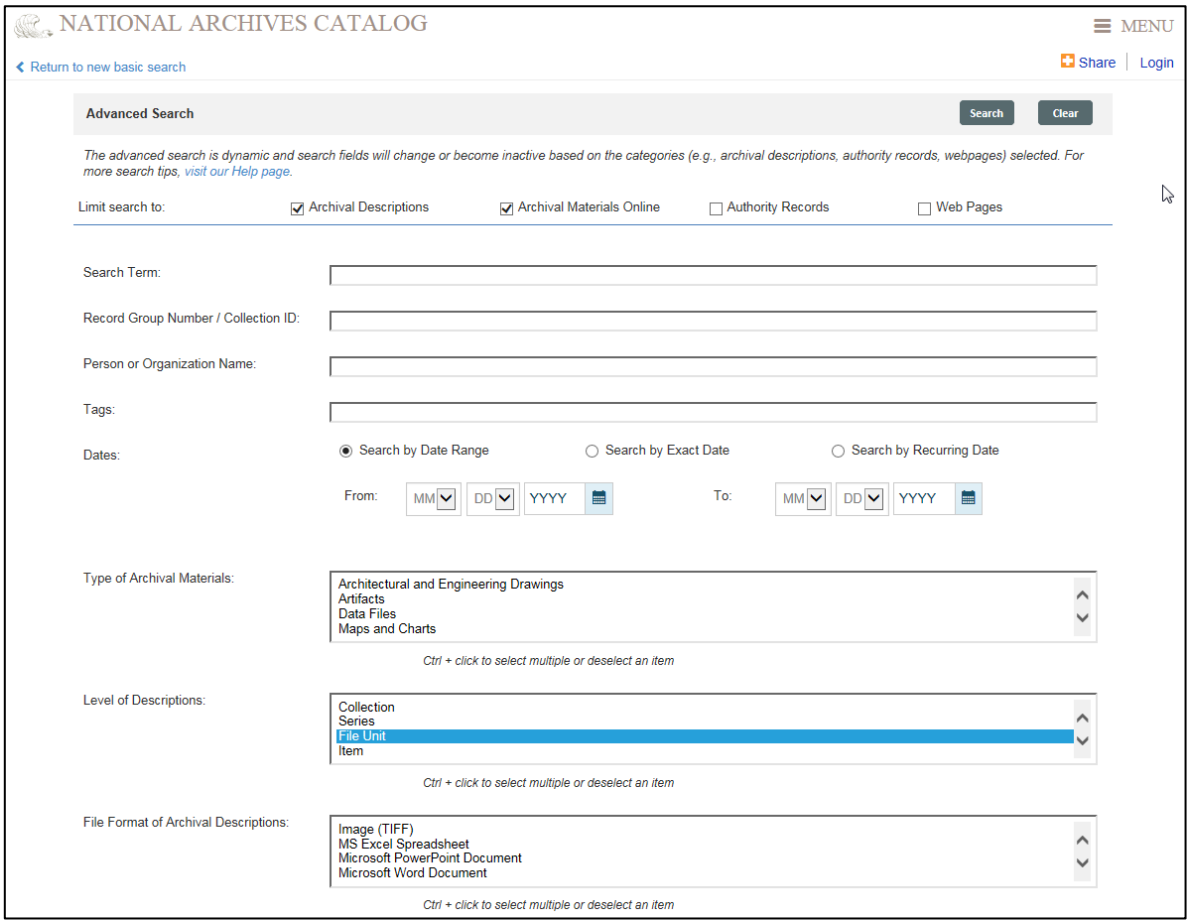


図 66 海外のデジタルアーカイブ事例(NARA Advanced Search 画面)

検索結果一覧には画像のサムネイルとタイトル、識別子等の資料の概要情報が表示され、画面左側にはファセットが表示され絞込みが行える(図 67)。

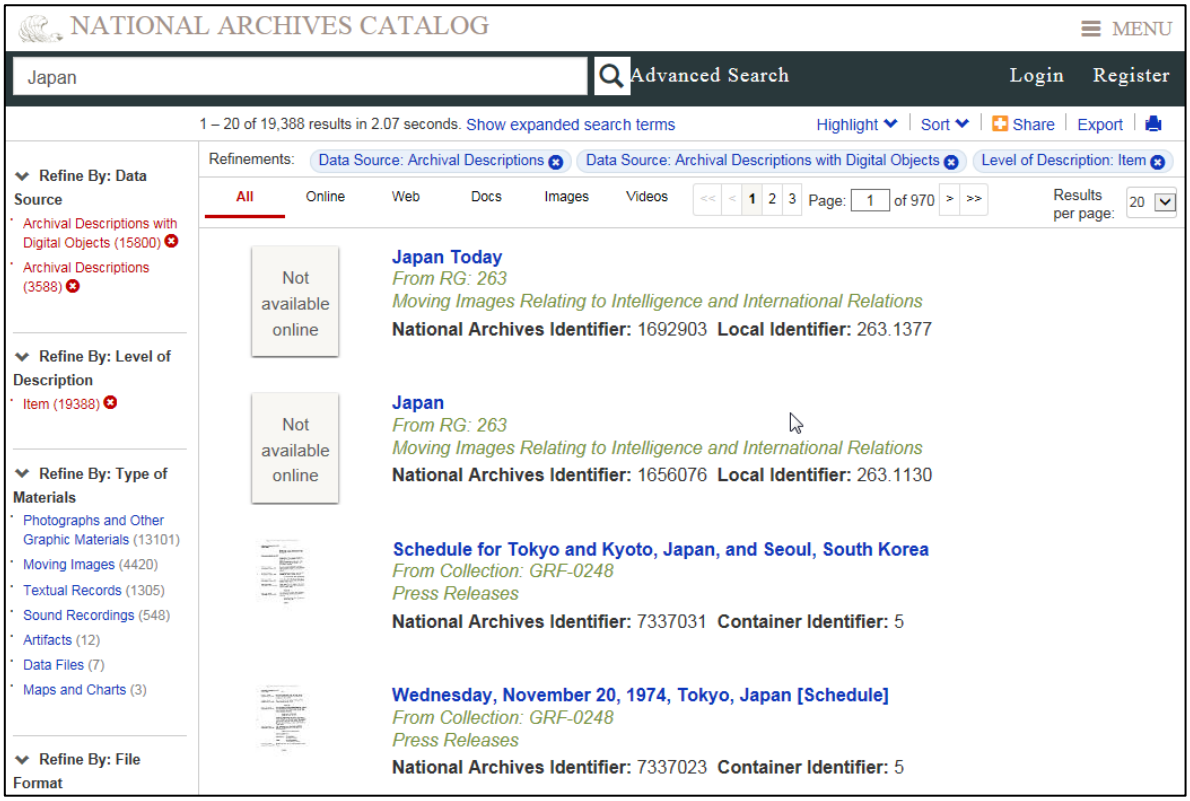


図 67 海外のデジタルアーカイブ事例(NARA 検索結果画面)

詳細表示画面には資料の詳細情報が表示されるとともに、タグやコメントを追加するエリアがある(図 68)。タグとコメントを追加するには、"community of citizen archivists"へのユーザー登録が必要である。"Share"をクリックすると、Facebook や Twitter の他、世界中の 200 近い SNS サービスを選択して連携することができる。

画面左上には Record Group、Series、File Unit、Item の階層が表示される。

図 68 海外のデジタルアーカイブ事例(NARA 詳細表示画面)

なお、NARA はホームページも含めてレスポンシブウェブデザイン(本報告書 2.2.10 項(2)(イ)参照)となっており、画面サイズに合わせて動的にレイアウトが変化するようになっている。

- TNA

TNA が提供する Discovery の検索機能として、トップページからのキーワード検索と、"Advanced search" が用意されている。"Advanced search"は、記録(Records)に対する単語やフレーズによる検索、除外する単語、レファレンス指定、日付の指定、作成者(Record creators)に対する種別を、3 段階のカテゴリから選択して検索が行える(図 69)。

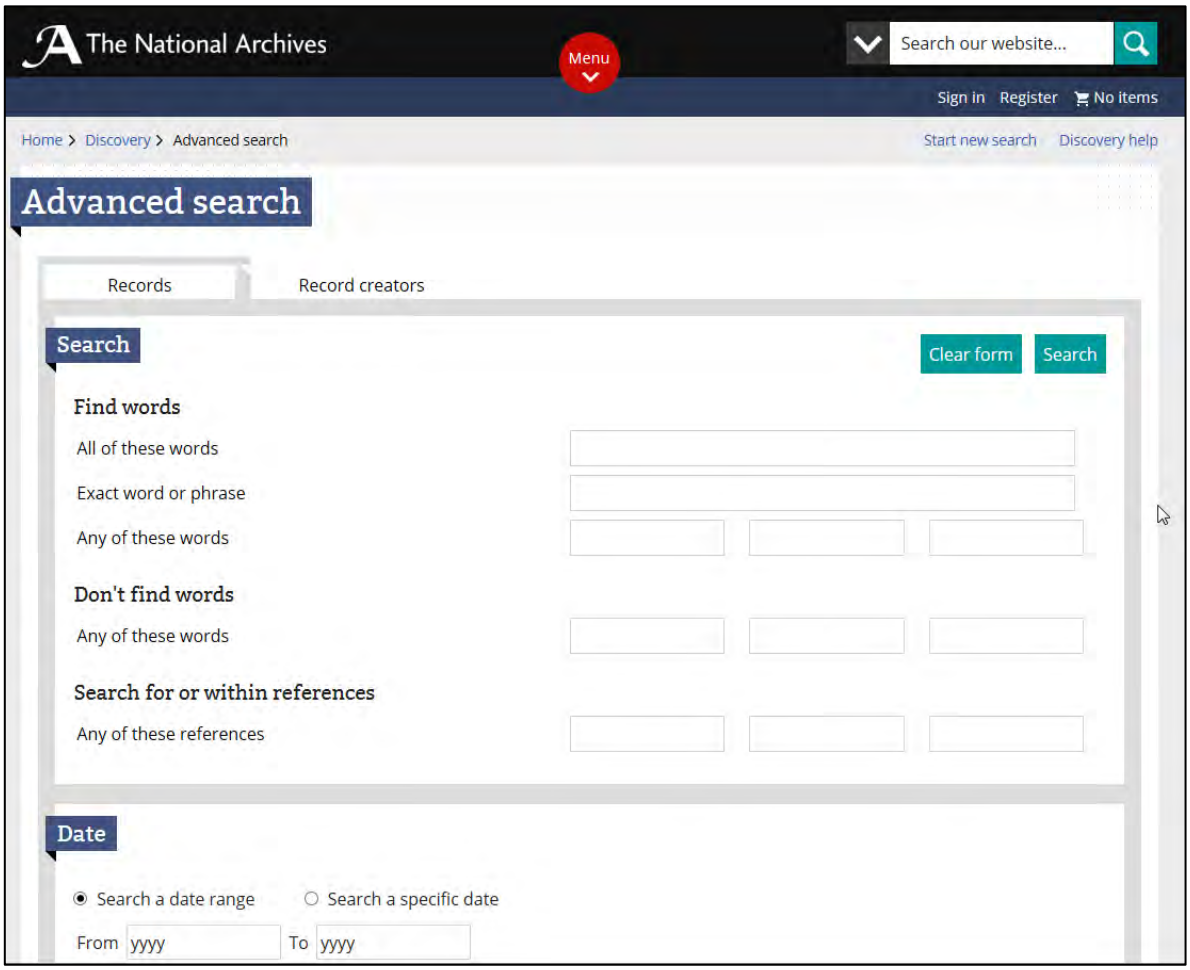


図 69 海外のデジタルアーカイブ事例(TNA Advanced search 画面)

検索結果一覧にはタイトル、所蔵機関、レファレンス等の資料の概要情報が表示され、画面左側にはファセットが表示され絞込みが行える(図 70)。検索結果は記録(Records)と作成者(Record creators)をタブにより切り替えられるようになっている。

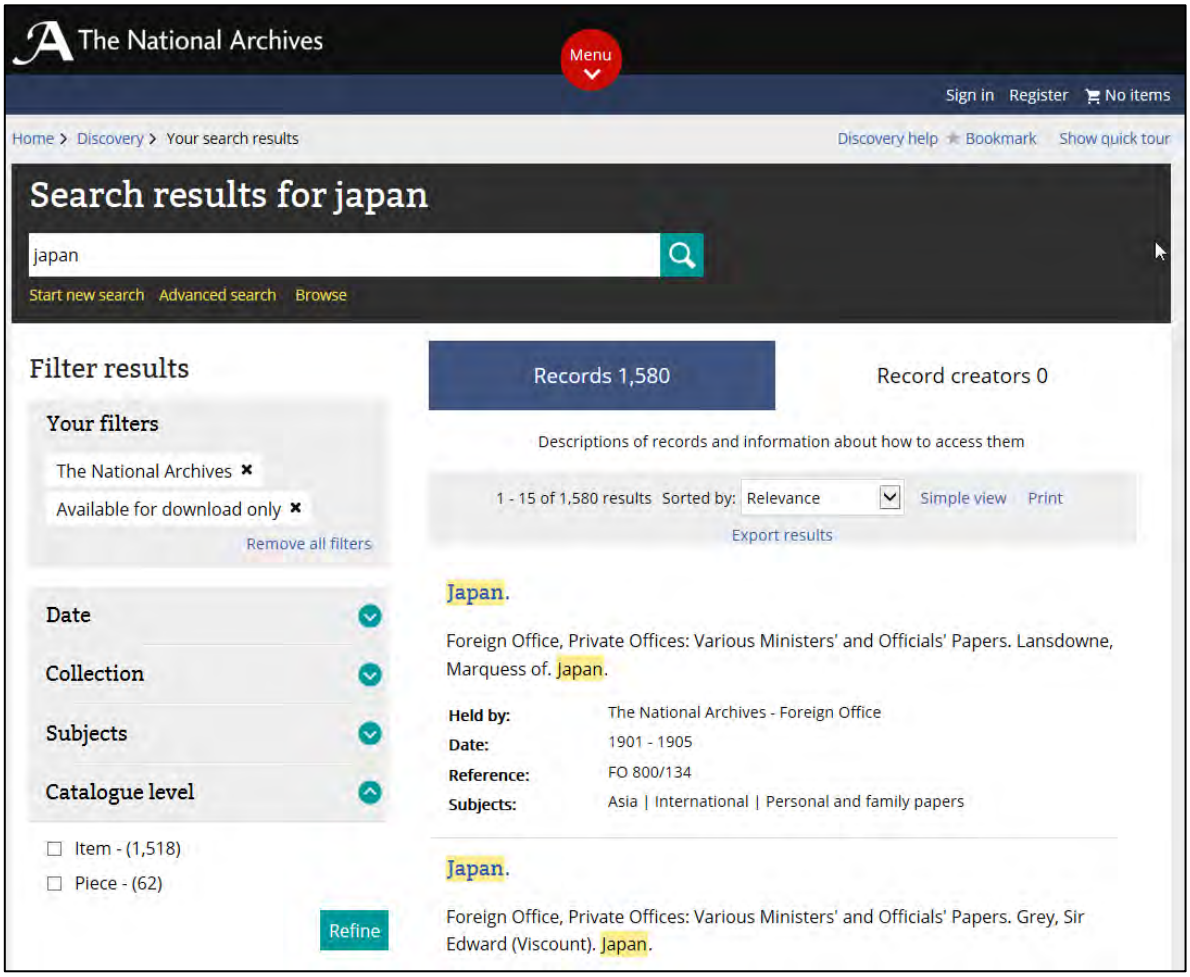


図 70 海外のデジタルアーカイブ事例(TNA 検索結果画面)

検索結果一覧で選択した資料の詳細情報を表示する詳細表示画面(図 71)では、デジタルコンテンツがある場合は"Add to basket"ボタン、"Show Images"リンクが表示される。画面右側にはタグを追加するエリア、内容の誤りを連絡するためのフォームエリアがある。タグを追加するにはユーザー登録が必要である。

画面下には"Context of this record"が表示され、部局と分類が階層的に表示される。

また、「＋」ボタンをクリックすると、Facebook や Twitter の他、世界中の 200 近い SNS サービスを選択して連携することができる。このユーザーインターフェースは NARA と同様であり、外部サービスを利用している。

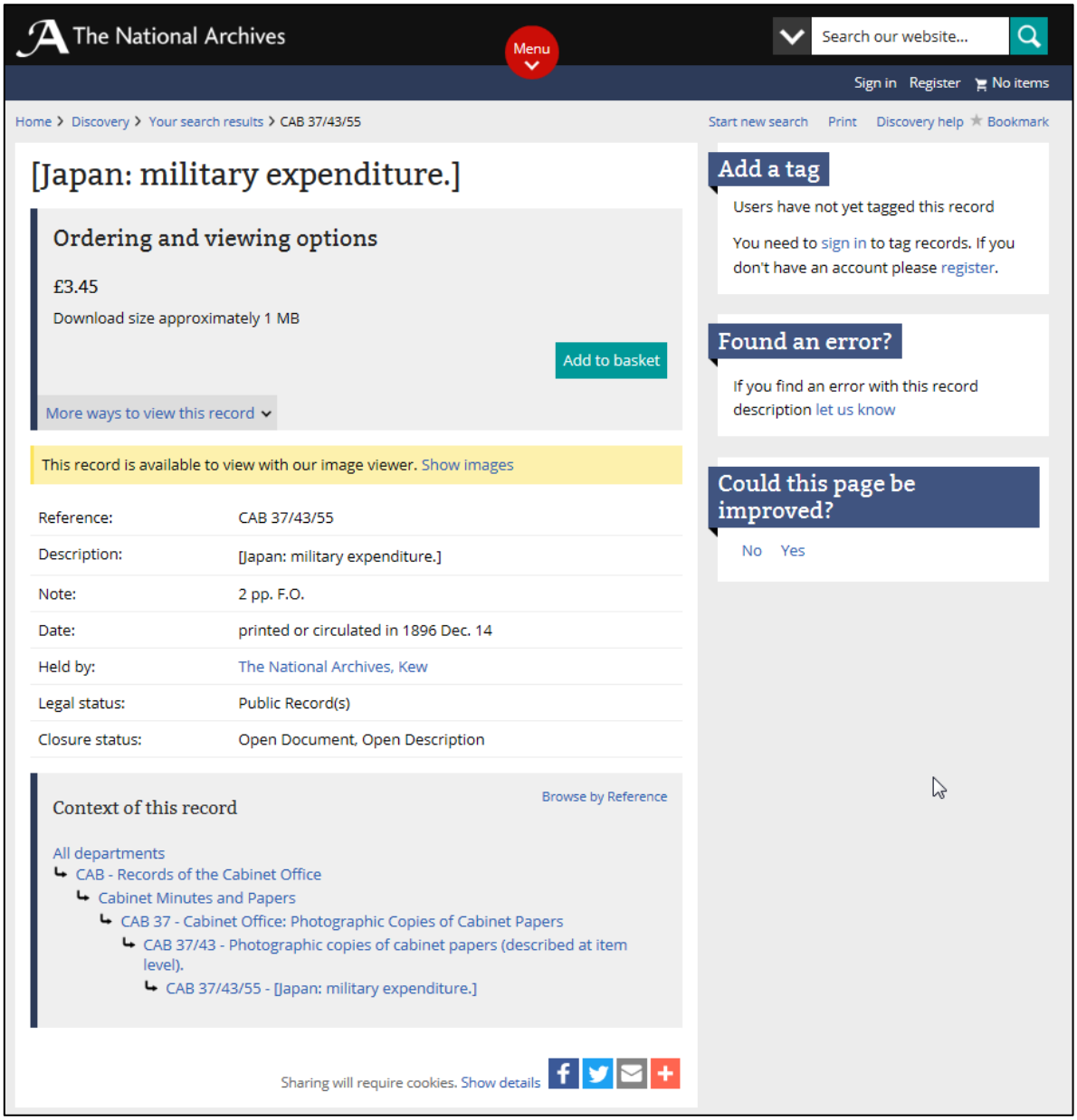


図 71 海外のデジタルアーカイブ事例(TNA 詳細表示画面)

”Context of this record”をクリックすると、行政機関の階層によるブラウズ画面が表示される(図 72)。ツリー型のユーザーインターフェースではないが、TNA 以外のアーカイブ名や作成者種別によるブラウズも行える。

図 72 海外のデジタルアーカイブ事例(TNA ブラウズ画面)

Discovery の特徴的な検索機能として、英国内外の 2500 以上の所蔵機関から、コレクションの作成者を元にアーカイブを検索する" Find an archive"という機能がある(図 73)。

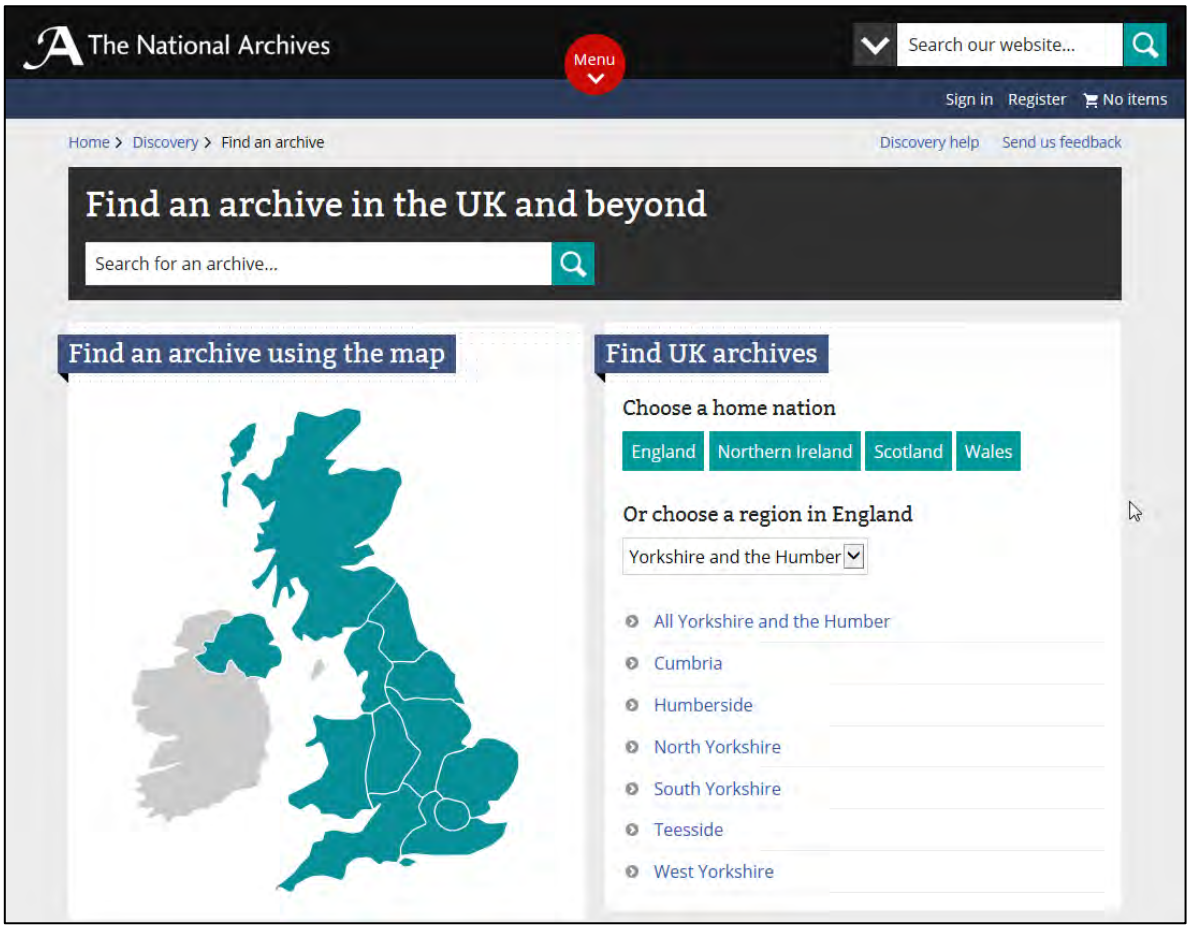


図 73 海外のデジタルアーカイブ事例(TNA Find an archive 画面)

TNA もホームページも含めてレスポンシブウェブデザインとなっており、画面サイズに合わせて動的にレイアウトが変化するようになっている。

- NAA

NAA が提供する RecordSearch の検索機能として、"Basic search"、"Advanced search"、"Name Search"、"Photo Search"、"Passengers arrivals index"が用意されている。"Basic search"はキーワードと日付のみの指定が行える。"Advanced search"は、Commonwealth Record Series (CRS) System と呼ばれる資料のレベルと機能、人名を最初に選択することで(図 74)、それに応じた検索項目を指定できる。例えば Item を選択すると、タイトル、シリーズ番号(Series number)、日付(Date)、物理的形式(Physical format)といった項目により検索が行える(図 75)。"Name Search"では、記録のカテゴリと姓(Family Name)による検索が、"Photo Search"ではキーワードと日付(Date)、主題(Subject)一覧による検索が、"Passengers arrivals index"では 1898 年から 1966 年までの船舶及び航空機の乗客名簿から、氏名、船名、到着地等による検索が行える。

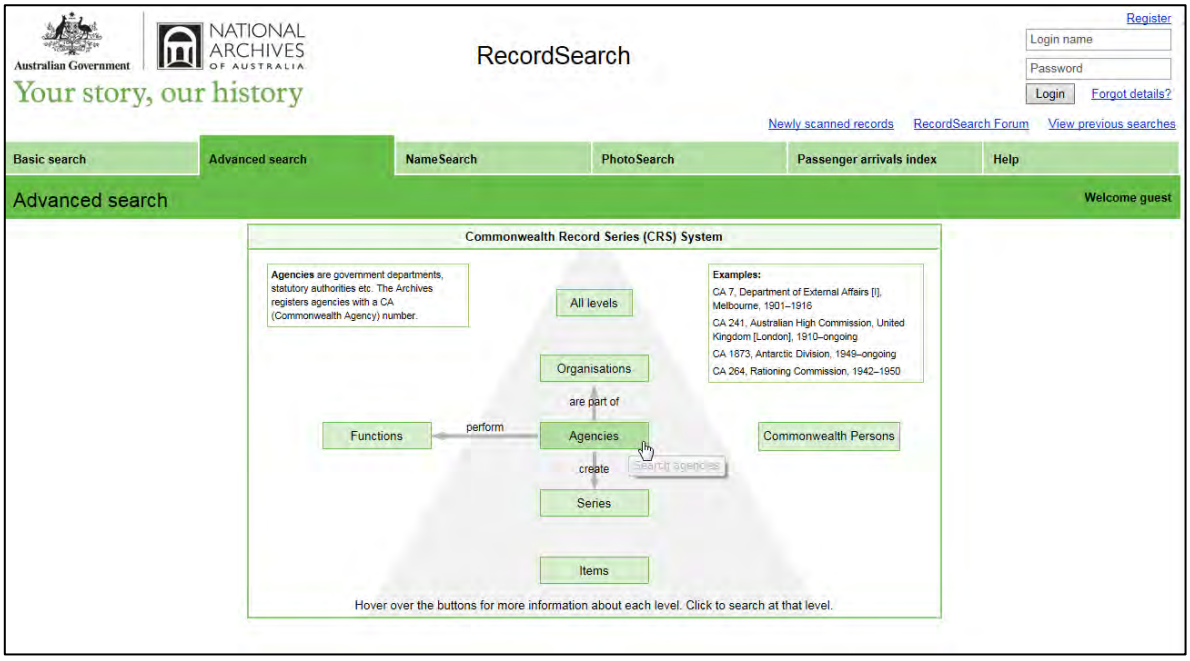


図 74 海外のデジタルアーカイブ事例(NAA Advanced search 画面)

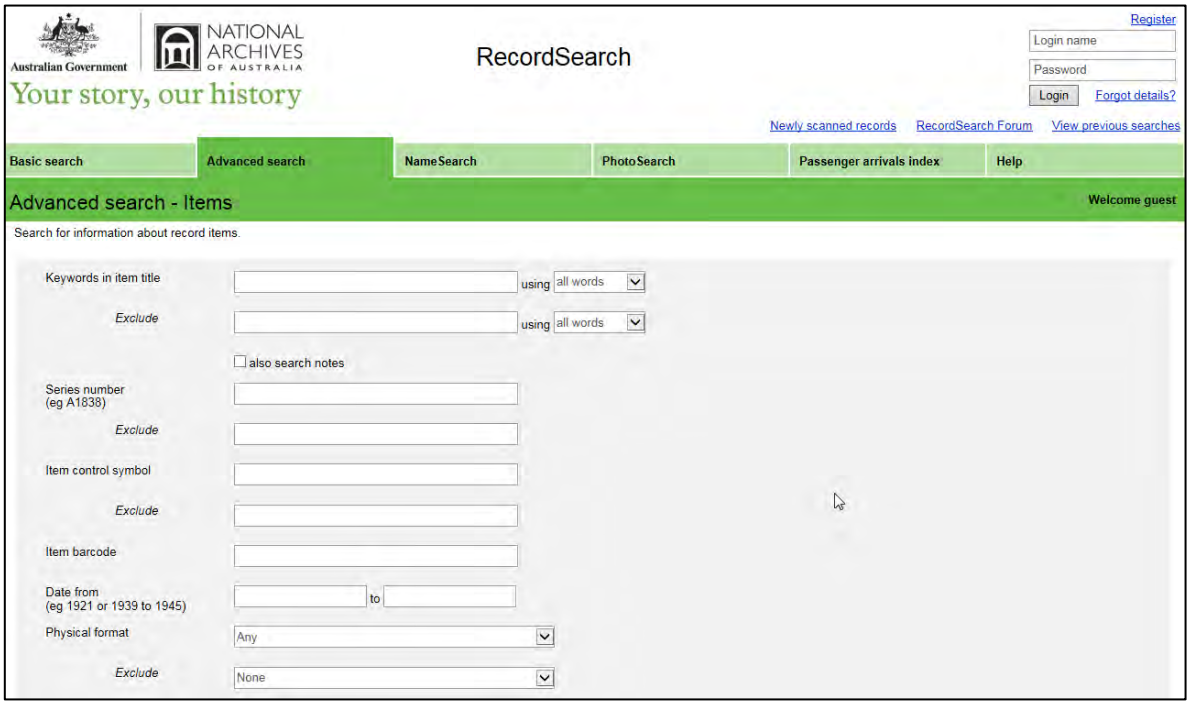


図 75 海外のデジタルアーカイブ事例(NAA Advanced search-Items 画面)

検索結果一覧の表示は最初の検索により内容が異なるが、Items を検索した場合はタイトル、シリーズ番号、日付、所蔵場所等の資料の概要情報が表示される(図 76)。

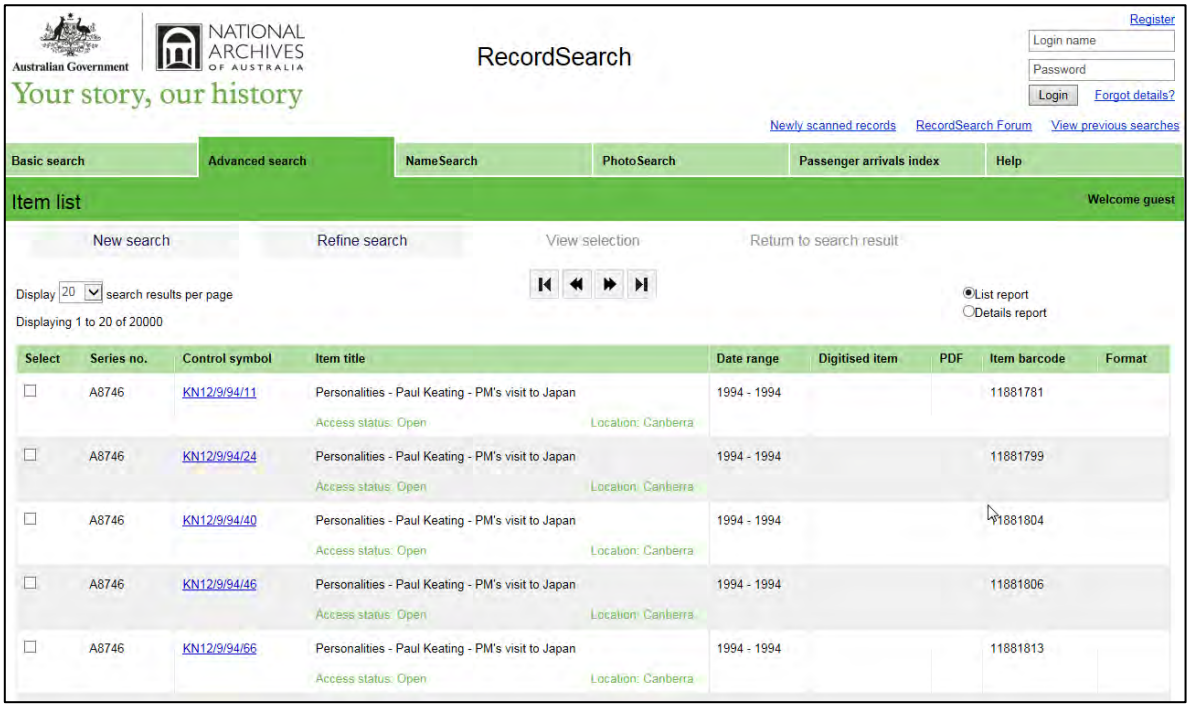


図 76 海外のデジタルアーカイブ事例(NAA 検索結果画面)

詳細表示画面には資料の詳細情報が表示されるとともに、閲覧可能なデジタルコンテンツがある場合は"View digital copy"リンクが表示される。デジタルコンテンツが無い場合は"Request Copy"リンクが表示される(図 77)。

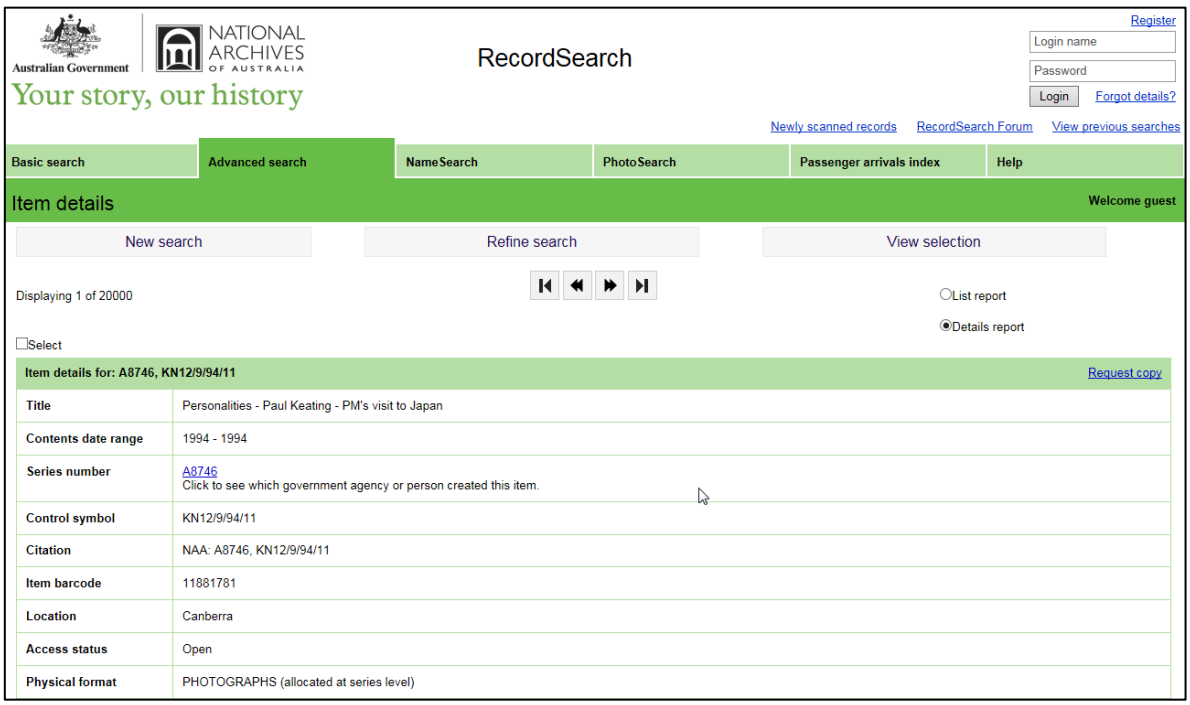


図 77 海外のデジタルアーカイブ事例(NAA 詳細表示画面)

NAA もホームページも含めてレスポンシブウェブデザインとなっており、画面サイズに合わせて動的にレイアウトが変化するようになっている。

（イ）画像の配信・閲覧機能

- NARA

NARA が提供する National Archives Catalog における画像の閲覧方法は、ポップアップ型(図 78)と画面埋め込み型(図 79)の 2 通りがある。ポップアップ型は検索結果一覧からサムネイルをクリックした場合に、検索結果一覧の上にポップアップで画像ビューアが表示される。画面埋め込み型は、詳細情報表示画面の上部にビューアが埋め込まれた形で表示される。

いずれの方式でもマウスホイールによる拡大・縮小、ドラッグによるスムーズな閲覧が行え、JPEG 形式でダウンロードが可能である。

画面埋め込み型の場合、“View/Add Contributions”ボタンをクリックすると画像に対するタグやコメントの追加が行える(詳細表示画面と同様にユーザー登録が必要である)。

マルチページの PDF 画像の閲覧や、動画・音声も再生可能である。

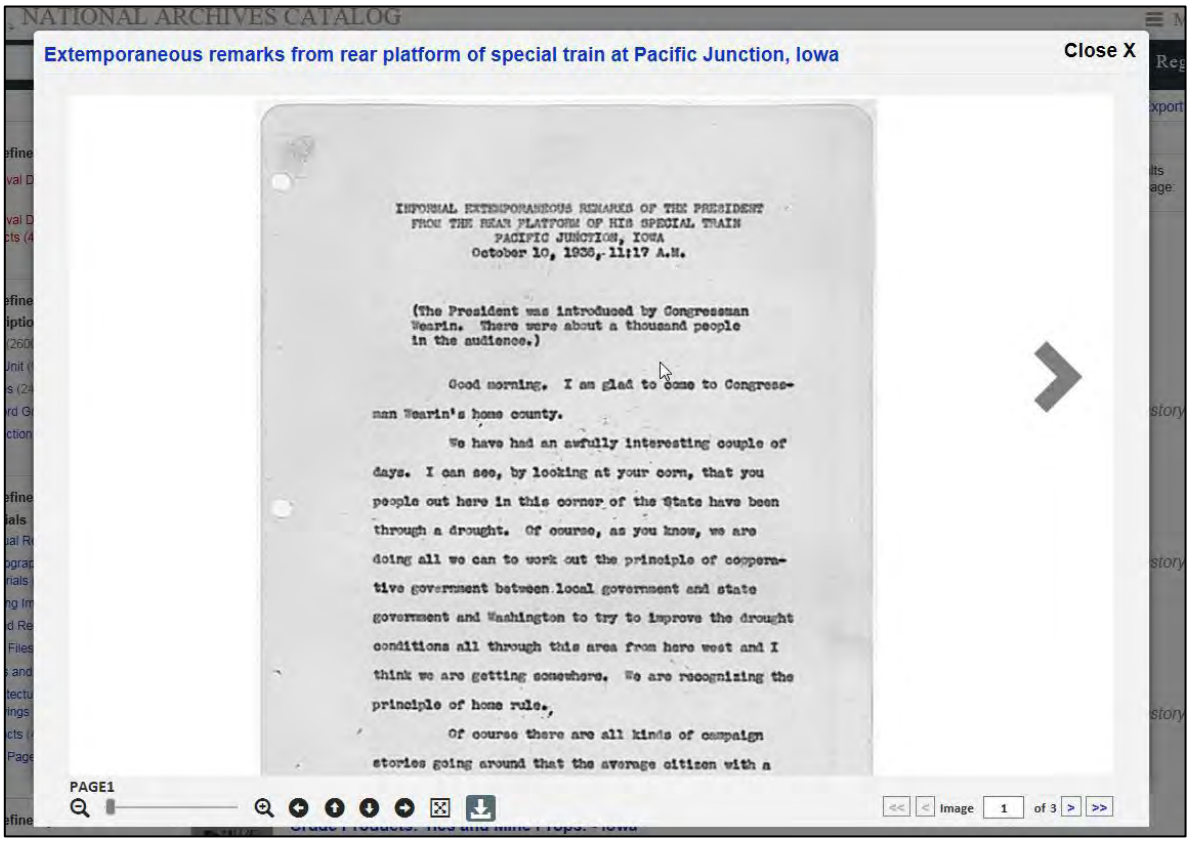


図 78 海外のデジタルアーカイブ事例(NARA 画像閲覧画面(ポップアップ型))

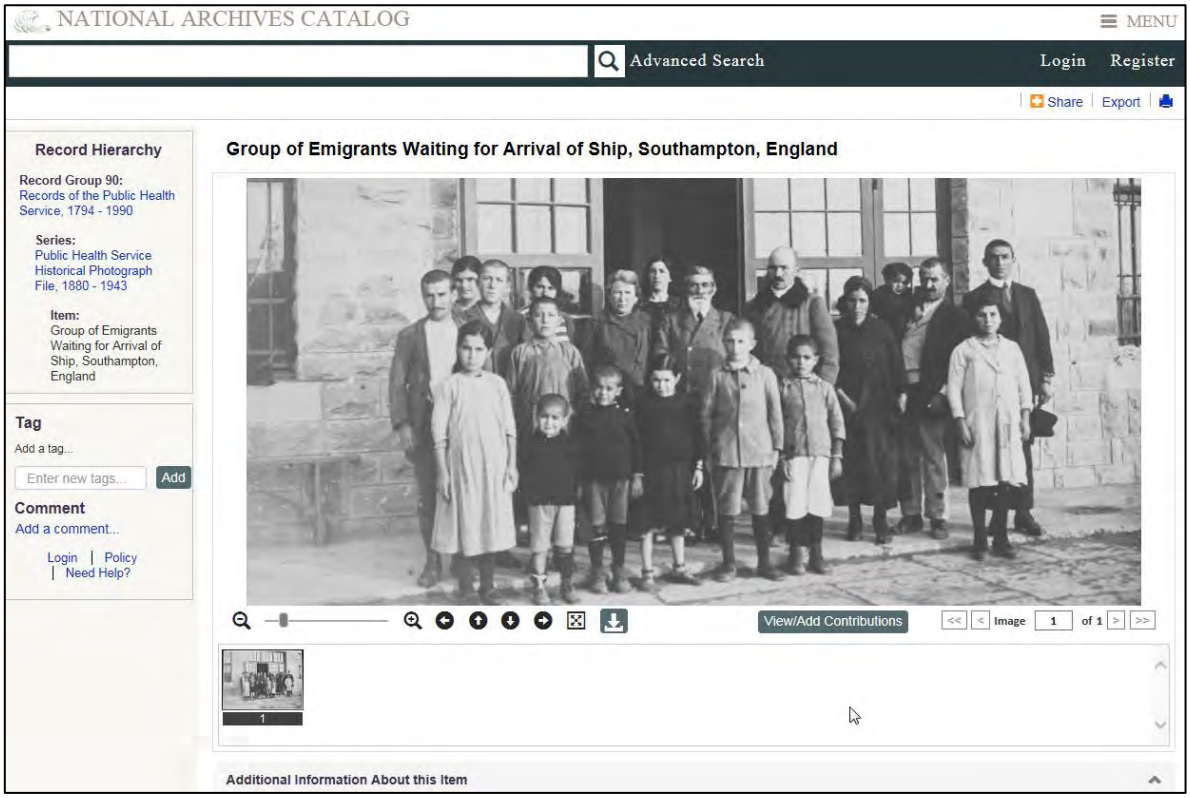


図 79 海外のデジタルアーカイブ事例(NARA 画像閲覧画面(画面埋め込み型))

- TNA

TNA が提供する Discovery における画像ビューアは画面埋め込み型であるが、全画面表示も可能である。マウスホイールによる拡大・縮小、ドラッグによるスムーズな閲覧が行える。

ただし、ビューアで閲覧可能な画像は低解像度でかつマスクがかけられた状態となっており(図 80)、高解像度の完全な画像を閲覧するには"Add to basket"でバスケットへ追加し、有償でダウンロードしなければならない。

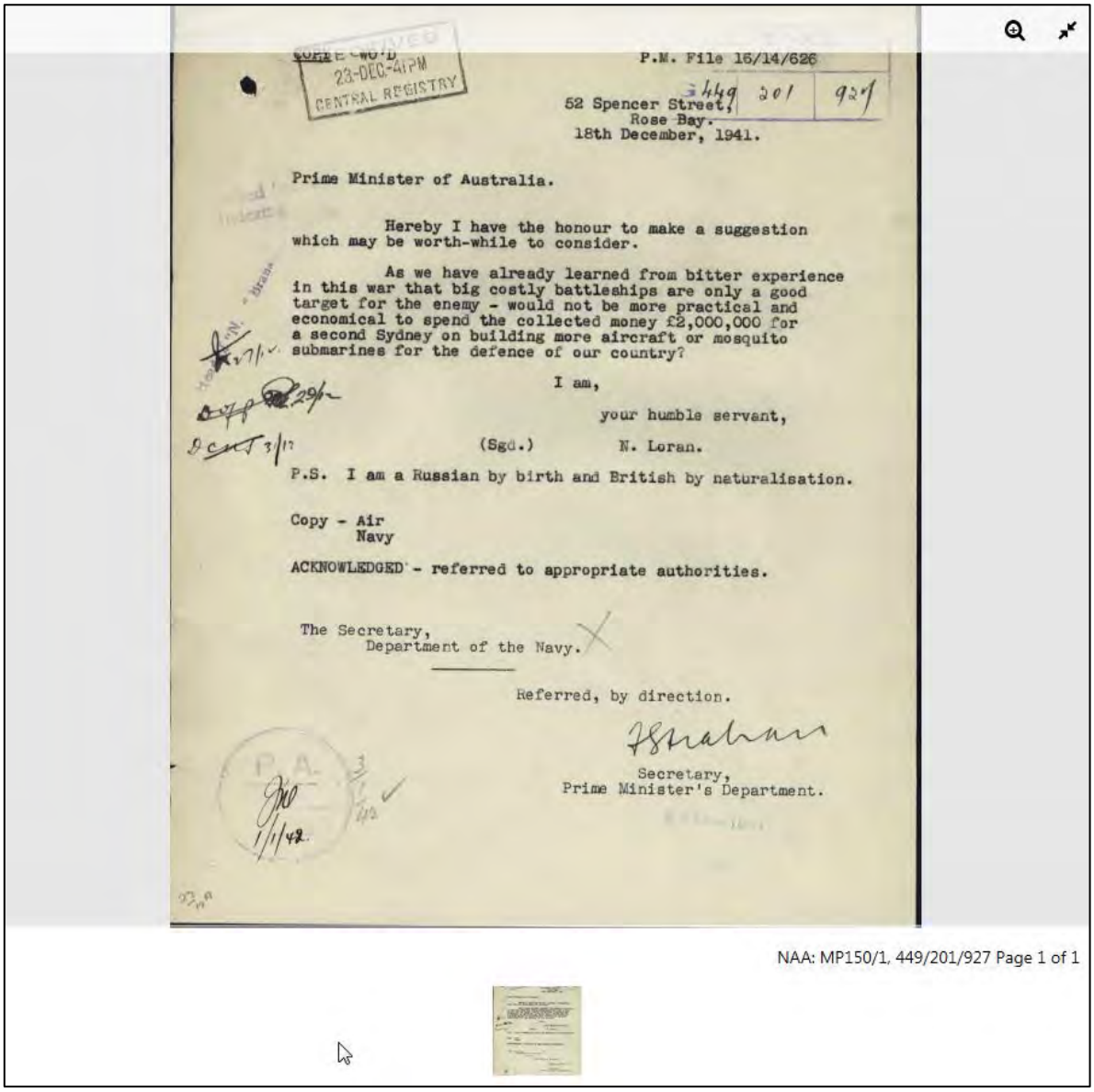
図 80 海外のデジタルアーカイブ事例(TNA 画像閲覧画面)

- NAA

NAA が提供する RecordSearch では、検索結果一覧からアイコンをクリックするか、詳細表示画面の”View digital copy”をクリックすると概要情報と全体画像が表示される(図 81)。この時点では画像の拡大・縮小や移動は行えないが、サムネイル一覧へ切り替えることができる。概要の下には SNS と連携するボタンが表示されているが、対象は Twitter、Facebook、Google+、メールの 4 種類に限定されている。

画像をクリックするとビューア画面が表示される(図 82)。拡大・縮小は画面右上のボタンで段階的に行えるのみで、ドラッグによる表示領域の移動は行えるが、NARA、NAA のビューアと比較すると操作性は劣る。

図 81 海外のデジタルアーカイブ事例(NAA 画像閲覧画面 1)



P.M. File 16/14/626

23-DEC-41 PM
CENTRAL REGISTRY

449 201 927

52 Spencer Street,
Rose Bay.
18th December, 1941.

Prime Minister of Australia.

Hereby I have the honour to make a suggestion which may be worth-while to consider.

As we have already learned from bitter experience in this war that big costly battleships are only a good target for the enemy - would not be more practical and economical to spend the collected money £2,000,000 for a second Sydney on building more aircraft or mosquito submarines for the defence of our country?

I am,

your humble servant,

(Sgd.) N. Loran.

P.S. I am a Russian by birth and British by naturalisation.

Copy - Air
Navy

ACKNOWLEDGED - referred to appropriate authorities.

The Secretary,
Department of the Navy.

Referred, by direction.

Secretary,
Prime Minister's Department.

NAA: MP150/1, 449/201/927 Page 1 of 1

図 82 海外のデジタルアーカイブ事例(NAA 画像閲覧画面 2)

- Europeana

Europeana が提供する Europeana Collections での画像閲覧は、画像をそのまま表示するものと、拡大・縮小機能を備えたものの 2 つが確認できる。対象機関が多く、収集されるコンテンツも大量となるため、元のコンテンツに応じた閲覧方法を提供しているものと思われる。

Europeana では画像だけでなく動画・音声も配信されている。

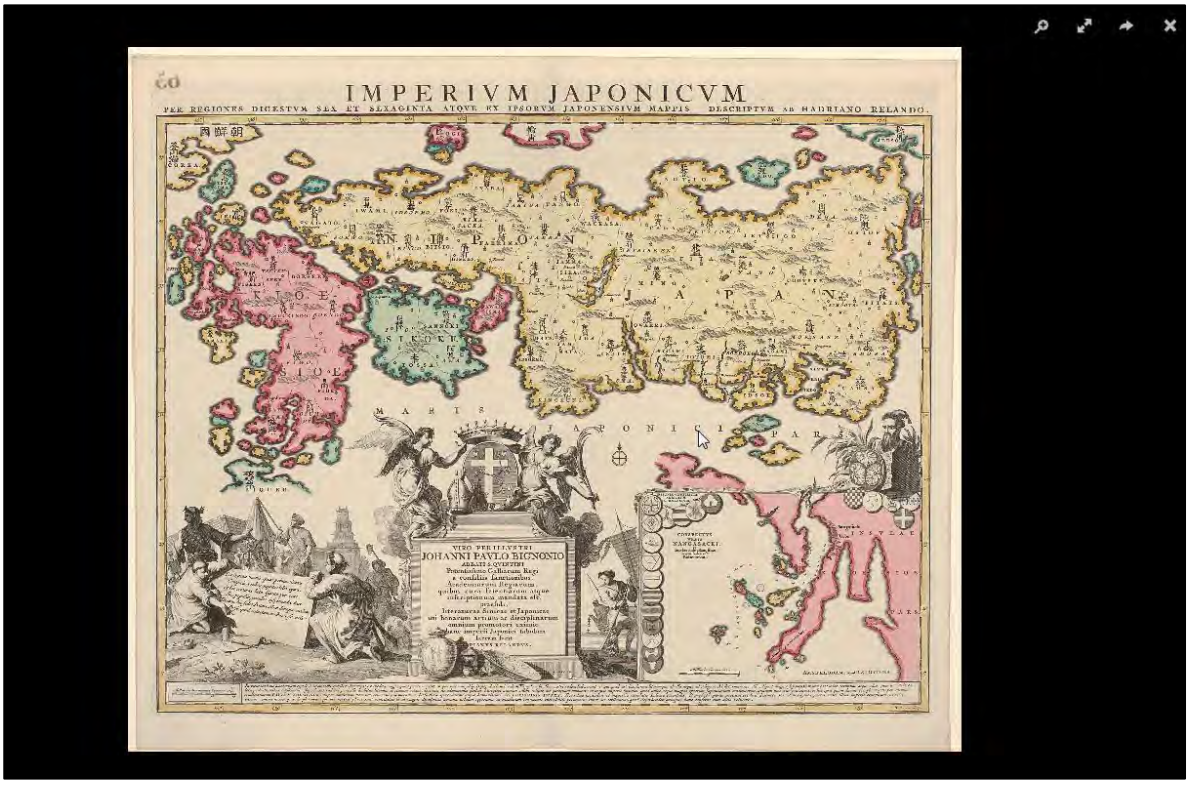


図 83 海外のデジタルアーカイブにおける画像ビューアの事例(Europeana)

（ウ）横断検索機能

NARA、NAA、TNA とも各国内の公文書管理機関の所蔵資料を対象とした検索が行えるようになっており、前述の検索機能と明確に区別された横断検索機能は見当たらなかった。

- Europeana

欧州連合の 100 を超える機関から、5,400 万件を超える美術品、工芸品、図書、動画、音声が検索できる。キーワード検索とコレクション、種別による検索が行え、検索結果からはファセットによる絞込みが行える(図 84)。ファセットはコレクション、ファイル種別、国、言語といった項目の他、再利用(Re-use)が可能なもの、限定的に再利用が可能なもの、再利用が不可のもので絞り込むことができる。

詳細情報には SNS 連携ボタンが配置されており、Facebook、Twitter、Pinterest、Google+、Tumbler の 5 種類の SNS との連携が可能である。

なお、Europeana の画面でもレスポンシブウェブデザインが採用されている。

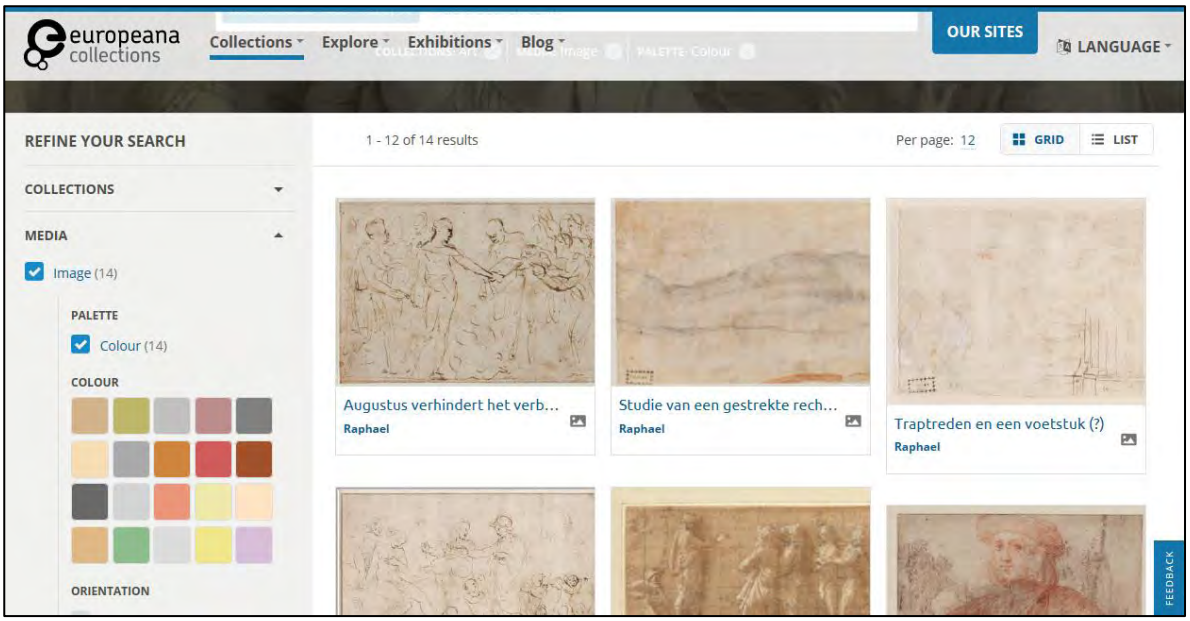


図 84 海外のデジタルアーカイブ事例(Europeana 検索結果画面)

（エ）考察

海外事例として 4 機関のデジタルアーカイブの調査を行ったが、各国・地域の事情により差異はあるものの、検索ではファセット機能が 3 機関で実装されており、2 機関で拡大・縮小や移動がスムーズに行える画像ビューアが実装されている。また、SNS 連携は対象数の差はあるが 4 機関とも実装されており、4 機関ともスマートデバイス向けのレスポンシブウェブデザインを採用している。

今回対象とした機関は国家または地域レベルでの大規模なデジタルアーカイブを構築している事例であるが、特にデジタルコンテンツの配信と利用者の利便性向上を主眼に置いているという共通点がある。NARA と TNA では利用者が使いやすい画像ビューアを採用しており、機能や操作性も共通点が多い。ファセット機能は「2.2.8 情報検索技術」で調査した全文検索エンジンの機能として備わっていることや、利用者がデータの内容に応じて簡単に絞込みが行えるという利便性の高さが採用の背景にあると考えられる。また、SNS 連携やレスポンシブウェブデザインの採用は、利用者の利用環境やニーズの変化を踏まえた流れである。

今後のデジタルアーカイブ・システムにおいては、単に目録検索や画像閲覧だけでなく、より利用者の視点に立ったサービス・機能が求められることになる。上記のファセット機能、SNS 連携、スマートデバイス対応(レスポンシブウェブデザイン)はデジタルアーカイブの標準機能として実装されることが望ましい。

高機能な画像ビューアについても同様に標準的に実装されることが望ましいが、デジタルアーカイブの領域では「2.2.9 デジタルコンテンツ配信技術」で調査した IIIF という画像の相互運用性を高めるフレームワークの標準化が進められており、利用者が使いやすい画像ビューアを自分で選ぶことも可能となる。その点では利便性向上のために有益な技術である。

3.1.4. デジタルアーカイブの基本機能(業務)

デジタルアーカイブ・システムで提供する業務側の基本機能の在り方を確認するため、国立公文書館デジタルアーカイブとアンケート対象機関の状況を元に業務側の基本機能提供について調査した。

調査概要

調査目的	業務側の基本機能の状況把握
調査対象	・国立公文書館デジタルアーカイブ ・アンケート対象機関
調査方法	・国立公文書館デジタルアーカイブの仕様による業務関連機能の調査 ・関連情報に対する文献調査(ウェブ調査を含む。)
調査内容	・目録データの管理・登録機能の状況 ・利用状況把握機能の状況 ・画像の管理・登録機能の状況 ・閲覧業務関連機能

(1) 目録データの管理・登録機能

(ア) 国立公文書館デジタルアーカイブ

目録データの管理・登録機能として、以下の機能が実装されている。

- EAD データ一括登録機能(EAD/XML 形式(EAD2002 に準拠)または CSV 形式による一括登録)
- EAD データ個別登録機能(資料群階層、簿冊・件名/細目)
- EAD データ一括編集機能
- EAD データと画像データ等連携機能
- 電子公文書データ連携機能(電子公文書管理システムからのデータ取り込み)

また、特定歴史公文書等の解説文等の属性情報を管理するピックアップコンテンツ管理機能がある。

目録データ等はデータベース(RDBMS)で管理されている。外字管理の機能は存在しない。

(イ) アンケート対象機関の状況

目録データの管理は、各機関独自の事情により実装されている機能(へのニーズ)には違いがあるが、一般的なデータベースを採用している機関は、Excel や CSV によるデータ登録が可能であり、その方法に大きな差異はないものと考えられる。

外字に関わる動向として、情報処理推進機構(IPA)が推進している「文字情報基盤整備事業」がある[117]。文字情報基盤整備事業は行政で用いられる人名漢字等約 6 万文字の漢字を整備するプロジェクトである。行政の実務で人名等の正確な表記のために実際に必要とされているにも関わらず標準化されていないために、「外字」という形で使用されている文字を再整理し、国際標準に基づいて使用できる様にすることを目的としている。デジタルアーカイブでは、現代の行政文書に関しては一定の利用効果が期待できるが、その他の資料については独自の外字が必要となるケースが想定されるため、同事業の成果を活用する際は、対象となる資料の特徴を踏まえた検討が必要である。

---

117 http://mojikiban.ipa.go.jp/ (2017 年 1 月 24 日アクセス)

(2)　画像の管理・登録機能

（ア）国立公文書館デジタルアーカイブ

保存用画像データとして JPEG2000(可逆データ)、提供用画像データとして JPEG2000(非可逆データ)及び PDF を扱えるシステムとなっている。

画像管理・登録に関わる機能として、以下の機能が実装されている。

- 画像データ品質検証機能(納入データの品質検証)
- 画像データ登録機能(個別登録と一括登録)
- 大判画像データ登録変換機能(高精細な大判画像データを効率的に配信するための形式に変換する)
- 電子公文書データ一括登録機能(電子公文書管理システムからのデータ取り込み)
- 利用用代替物データ一括登録機能

（イ）アンケート対象機関の状況

画像データの管理は、各機関独自の事情により実装されている機能(へのニーズ)には違いがあるものと考えられる。なお、原資料の閲覧に際しての個人情報保護・プライバシー保護は、各機関独自の基準で行われている。

画像の無断利用（転載）を防ぐための「すかし」の付与や印刷制限などは、民間のサービスには存在するが、そのためのコストが必要となる。基本的に「市民への広い公開」を目的とするデジタルアーカイブでは、非公開事由以外のダウンロード制限を設けていない。しかし、個人情報・プライバシー保護の観点から閲覧制限が必要な資料については、利用規約などに権利関係を利用者が明確に理解できるよう記載するといった配慮が必要である。

(3)　利用状況把握機能

（ア）国立公文書館デジタルアーカイブ

利用状況把握には、アクセスログや検索キーワードのログ情報、利用請求に関する情報の収集及び解析機能が実装されている。また、アクセス状況の把握には外部サービス(Google Analytics)が利用されている。

他の利用実態把握に関する機能として、ウェブアンケート実施に関連した機能が実装されている。同機能を構成する主な要素を以下に記す。

- アンケート作成・保存・管理機能(アンケートフォームを職員が作成できる)
- アンケート公開機能(作成したアンケートフォームを一般利用者に公開する)
- アンケート集計機能(利用者の回答内容を集計する)

また、上記機能により作成されたアンケートフォームへの回答は、HTTPS により暗号化された通信経路を通して送信されるようになっている。

（イ）アンケート対象機関の状況

システムへのアクセス、とくに利用者による目録検索や画像閲覧回数といった利用状況の把握は、一般的ツールを用いた Web サーバのアクセスログ解析により可能である。

Google Analytics は高機能なアクセス解析が行えるサービスであり、1 カ月あたり 1,000 万ヒットまでは

無償で利用できる。ただし、外部(Google)へ IP アドレスや利用者がアクセスした URL 等の情報、利用者の Web ブラウザに保存された Cookie と呼ばれる情報が渡るためセキュリティポリシーにより利用できない可能性があり、企業が運営するサービスであるためサービス仕様の変更やサービス停止がありうる点に注意が必要である。外部サービスを利用する場合でも、最低限 Web サーバのアクセスログ保全は必須と言える。

また、アクセスログによる利用状況把握は、定量的な評価の指標としては役に立つが、利用者がどのようにデジタルアーカイブを利用しているか、どのような資料や機能を求めているか、操作性や利便性に満足しているかといった定性的な情報は把握できない。このような情報を把握するためには、利用者に対してアンケートを実施し、結果を分析するという方法がある。

アンケートは各機関の閲覧室等に用紙を置き、利用者に回答を記述してもらう方式もあるが、結果のデータ化や集計に手間がかかる。Web 上でアンケートを作成し、回答を求める方法は集計等の手間が最小限で済むが、システム構築のコストや回答に個人情報等が含まれる場合にセキュリティ上の配慮が必要となる。

(4) 閲覧業務関連機能

(ア) 国立公文書館デジタルアーカイブ

機関内で資料閲覧する利用者が、検索結果から利用請求書等を出力・印刷するとともに、データベース上に請求情報を格納して管理する機能として、以下の機能が実装されている。

- 利用請求業務管理機能(利用請求情報の更新・変更)
- 利用請求票管理機能(利用請求データの参照・情報追加)
- 利用請求データ検索・出力機能(条件指定による利用請求データの検索・CSV 形式での出力)
- 閲覧室利用者管理機能(閲覧利用者の登録・変更・削除)
- 利用請求統計機能(利用種別、利用方法等による帳票出力)

インターネットからの一般利用者の場合は、利用請求書等の出力のみ可能であり、その場合は上記の管理機能の対象にはならない。

(イ) アンケート対象機関の状況

検索システム上で利用請求書や複写申込書を出力できるシステムの事例が見られる。システムが無い、またはシステムと連動して出力ができない場合でも、多くの機関で利用のための申請フォーマットが用意されている。

請求情報の管理、たとえば資料種別ごと、月別の利用統計の調査などは各機関の業務に関わる部分であるが、システム化されておらず紙ベースでの業務となり、職員の負担増となっている可能性がある。アンケート結果で利用請求記録の管理機能に対するニーズが 8 割近くに上ったことからも、利用請求の管理はデジタルアーカイブ・システムにおいて重要な機能であるといえるが、利用請求に必要な項目や、管理すべき情報、データ化の方法について検討が必要である。

3.1.5. 利用者拡大のための機能・サービス

デジタルアーカイブ・システムで提供する利用者拡大のための機能・サービスの在り方を検討するため、全国の公文書館等における、デジタルアーカイブ・システムやホームページなど、インターネットを通じた機能・サービス提供の状況を調査した。

調査概要

調査目的	利用者拡大のための機能・サービスの状況把握
調査対象	・国立公文書館デジタルアーカイブ ・アジア歴史資料センター ・国立公文書館のホームページにおいて、国内外や類縁機関等、国立公文書館に関係のあるウェブサイトを集めたリンク集に示されている、国内109 機関 ・国内の公文書館類縁機関 ・海外のデジタルアーカイブ（NARA, Europeana）
調査方法	・上記機関のホームページ等の利用(平成 28 年 12 月 19 日現在)
調査内容	・利用者による情報発信支援の状況 ・利用者によるシステムへの能動的参加の状況 ・デジタルコンテンツの活用の状況 ・他の機関との連携の状況 ・その他特徴的な事例

(1)　利用者による情報発信支援

SNS(Facebook、Twitter 等)上で、デジタルアーカイブ・システムのコンテンツ(目録情報や画像)の URL を利用者が発信することを支援する機能や、他の SNS 連携・活用事例について調査し、利用者拡大に資する在り方を検討した。

(ア) SNS 連携

公文書館等において目録詳細や画像に SNS(Facebook、Twitter 等)と連携するためのボタンを設置している事例があった(図 85)。ツイートボタンをクリックすることで、URL 等入力済みの投稿画面がポップアップで表示される。

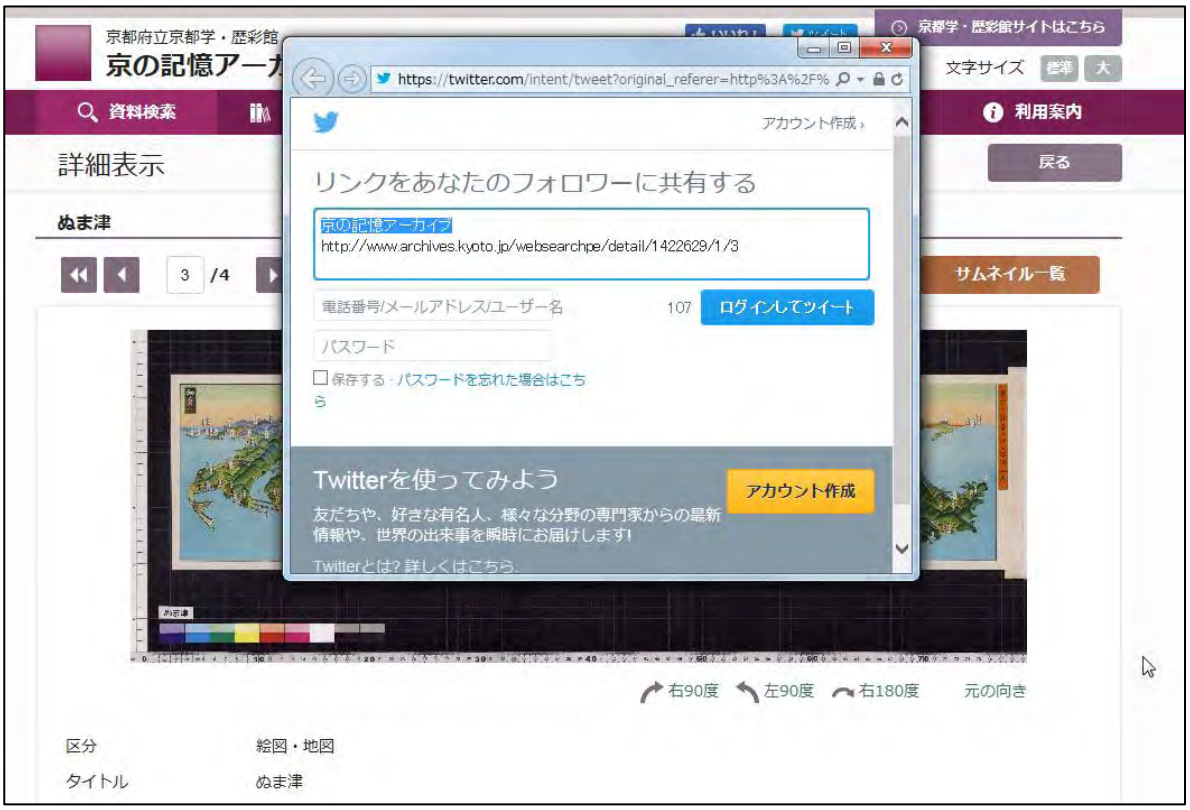


図 85 京都府立京都学・歴彩館「京の記憶アーカイブ」における Twitter との連携例

(イ) ソーシャルタギング

ブックマークした Web サイトに利用者が任意の分類キーワードを付け、Web 上で共有できる仕組みである。海外では Delicious 等のサービスがあり、国内でもはてなブックマークといったサービスが提供されている。

ソーシャルタギングは利用者が内容を理解し、適切なタグを付与することで有効な情報検索手段となりうるが、タグとして用いられるキーワードは利用者ごとに異なり統制されていないため、利用者を誤った結果へ導き混乱を招く懸念がある。

(ウ) 考察

目録や画像の詳細に SNS と連携するためのボタンを設置している事例や、各機関の広報機能の一つ

として SNS を活用(Facebook 上で公式ページを用意し、資料の紹介を行う等)している事例があり、利用者による二次的な広報を期待できることから、機関によっては広報手段として SNS を積極的に採用している。しかし、昨今の SNS による不適切な情報発信と、予想しない非難を受ける(いわゆる「炎上」)状態の発生の増加を受けて、SNS 利用のガイドラインを定めて運用する自治体が増えており、情報発信のたびに承認を得る手続きが必要となる場合もある。利用者がデータ登録や情報発信に寄与する機能は利用者拡大に有効と考えられるが、運用方針の明確化が必要である。

利用者による情報発信を支援するための機能は、該当資料にリンクを搭載するだけでなく、「利用者が利用したい」と思わせるためのツールの充実が必要である。したがって、SNS の活用だけでなく、「レファレンスツール」の充実とあわせて検討することが必要である。

(2) 利用者によるシステムへの能動的参加

(ア) 最新動向

市民協働の重要性が認識され、各機関においては、整理作業(目録作成作業)に市民ボランティアを導入している事例も見られる(尼崎市立地域研究史料館、福井県文書館など)。尼崎市立地域研究史料館においては、館内限定で広報課撮影写真等を公開し、市民からの情報提供を中心とした写真データベースを構築しつつある。特に地域に残された写真整理の際には、利用者(市民)の「記憶」が重要なキーとなる場合がある。したがって、画像をデジタルアーカイブ上にアップし、メタデータ付与に利用者(市民)が参画するといったことは、運用上の課題はあるとはいえ十分検討に値すると考えられる。

データの品質や責任を明確にしたうえで、近年広まりつつあるクラウドソーシングを活用することも低コストでデータ化を行う選択肢としては有効と考えられる。

（イ）事例

- アジア歴史資料センター

アジア歴史資料センターにおいては、TOP ページのお問い合わせリンクから目録データや誤字脱字といった不具合情報の受付フォーム(図 86)が用意され、寄せられた不具合情報からデータベースの質的向上のために使用している。

図 86 アジア歴史資料センターの不具合情報（誤字脱字を含む）フォーム[118]

118 https://www.jacar.archives.go.jp/aj/enqform/trbform.do （2017 年 1 月 24 日アクセス）

- Europeana

Europeana では、データの品質に係るガイドラインや利用規約を整備のうえ、「Europeana 1914-1918」等のコレクションにおいて、ユーザー参加型コンテンツの促進を図っている。投稿されたメタデータは各機関のアーキビストがチェックする仕組みとなっている。

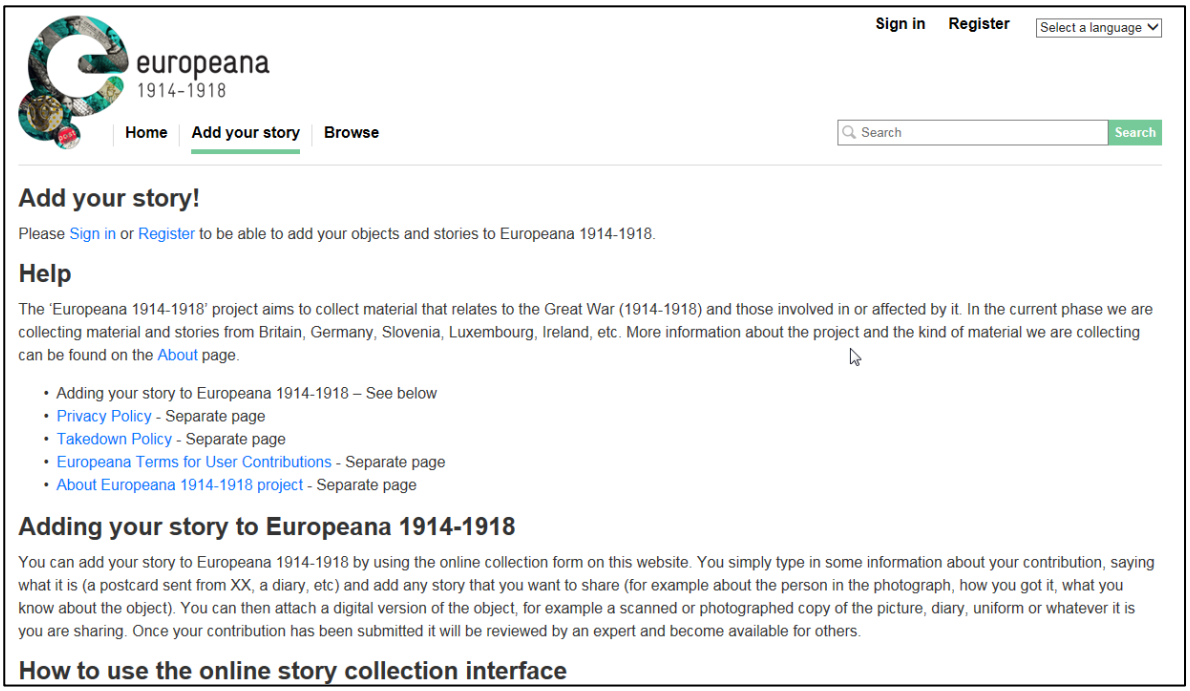


図 87「Europeana 1914-1918」の Add your story ページ

- NARA(National Archives and Records Administration) Citizen Archivist Dashboard[119]

市民アーキビストの参加の場として、2011 年 12 月に開設された。利用者登録はメールアドレスのみであり、キーワードの制約を除いて難しい規則は無く、一般市民が参加しやすい。①タグ付け（"Tag Images and records"、資料目録に検索用のキーワードを付与する作業）、②文字起こし（"Transcribe Historical Documents"）、③オンラインレファレンス（"History Hub"）、④アップロード＆シェア（"Upload & share"、利用者がデジタルカメラで撮影した画像をグループで共有する）という流れで作業を行うことが可能となっている。

図 88 NARA Citizen Archivist Dashboard

(3) デジタルコンテンツの活用

(ア) デジタル展示によるコンテンツ活用

公文書館等においては、ホームページ上でデジタル展示や特別展などの形でデジタルコンテンツを公開している事例がある。

一方で、類縁機関である博物館においては、工芸品や標本等の収蔵資料をマルチタッチモニターで様々な展示物を直感的に操作できる仕組みを導入している事例[120]がある。

民間におけるデジタルサイネージの分野では、人の集まるターミナル駅の広告媒体や、周辺案内へ利用が進んでいる。

---

119 https://www.archives.gov/citizen-archivist （2017 年 1 月 24 日アクセス）

120 2012 年に国立民族学博物館において、ATR Creative 社のイメージファインダーを用いた展示が設置された。
http://www.minpaku.ac.jp/museum/enews/131otakara （2017 年 1 月 24 日アクセス）
http://www.atr-c.jp/showcase/ （2017 年 1 月 24 日アクセス）

（イ）SNS によるコンテンツ活用

公文書館等において Twitter や Facebook で情報発信を行っている事例があった(図 89、図 90)。投稿内容は、イベントの告知や開館時間の案内といった広報目的の他に、新たに公開した資料や時事に関連する資料を紹介している。Facebook では、ビジター投稿を利用し、利用者とコメントのやりとりをしている事例もあった。

図 89 東京都公文書館の Twitter 公式アカウント

図 90 尼崎市立地域研究史料館の Facebook 公式アカウント

（ウ）考察

デジタルコンテンツの活用においては、インターネット上で多くの利用者や潜在的利用者へ情報提供する形態がある一方、原資料を複製もしくは模造したもの、あるいは電子的に構成された展示物として、単体で物理的に展示する形態が採用されている。元来無体物であるデジタルコンテンツは様々な形態により展示へ活用することが可能であると考えられる。

また、展示資料としてだけではなく、デジタルサイネージ等の技術により、不特定多数の方へ向けた広告媒体としての活用も可能ではないかと思われる。デジタルコンテンツ等資料情報の展示等への二次利用は利用者拡大に有効であると考えられる。

SNS におけるデジタルコンテンツの活用についても、Twitter 等での情報発信に新規公開資料等に関するコンテンツを扱っている公文書館等の例もあることから、公文書館等において、SNS の活用を意識している機関があることが確認された。SNS は一次的に発信された情報の二次的利用が当たり前のように行われているメディアであり、デジタルコンテンツを SNS での情報発信に活用することにより、広範な情報伝播を期待することができるものと考えられる。

（4） 他機関との連携

（ア）国内の事例

国内の公文書館等における他機関との連携事例について調査を行った。

- 岡山県立記録資料館(岡山県立図書館との連携)

  岡山県立図書館の提供する電子図書館「デジタル岡山大百科」において、「郷土情報ネットワーク」の対象データベースの一つとして岡山県立記録資料館の所蔵データが検索できる。連携は岡山県立図書館が岡山県立記録資料館の所蔵データを OAI-PMH により収集する形で行われている。

- 秋田県公文書館(秋田県立図書館、秋田県立近代美術館、秋田県立博物館、秋田県埋蔵文化財センター、あきた文学資料館、秋田県生涯学習センターとの連携)

  平成 25 年 3 月よりデジタルアーカイブを公開しているが、県内の社会教育機関等 7 機関の所蔵データを横断検索することが可能となっている[121]。

  7 機関がそれぞれ異なる URL で検索画面を持っているが、基本的なデザインや機能は共通となっており、参加機関が共通のシステムを利用しているケースと推測される。そのため、参加機関の横断検索が容易に実現できていると考えられる。

- アジア歴史資料センター(琉球大学附属図書館との連携)

  アジア歴史資料センター(アジ歴)と琉球大学附属図書館は 2013 年 5 月に資料公開についての連携協定を締結し[122]、アジア歴史資料センターのデータベースから琉球大学附属図書館の一部資料を検索できるようになっている。対象は「宮良殿内文庫」と「矢内原忠雄文庫植民地関係史料」であるが、今後拡大する予定となっている。

  連携は琉球大附属図書館から定期的にデータをアジ歴のデータベースに取り込む形で行われている。

(イ) 海外事例

先に記載した通り、Europeana では OAI-PMH により参加機関のメタデータの収集を行っている。多くのデータ提供者とアグリゲータが介在するため、収集するメタデータの集合であるデータセットに対するメタデータ(データ要素)として Europeana Dataset Profile を定めて公開している[123]。

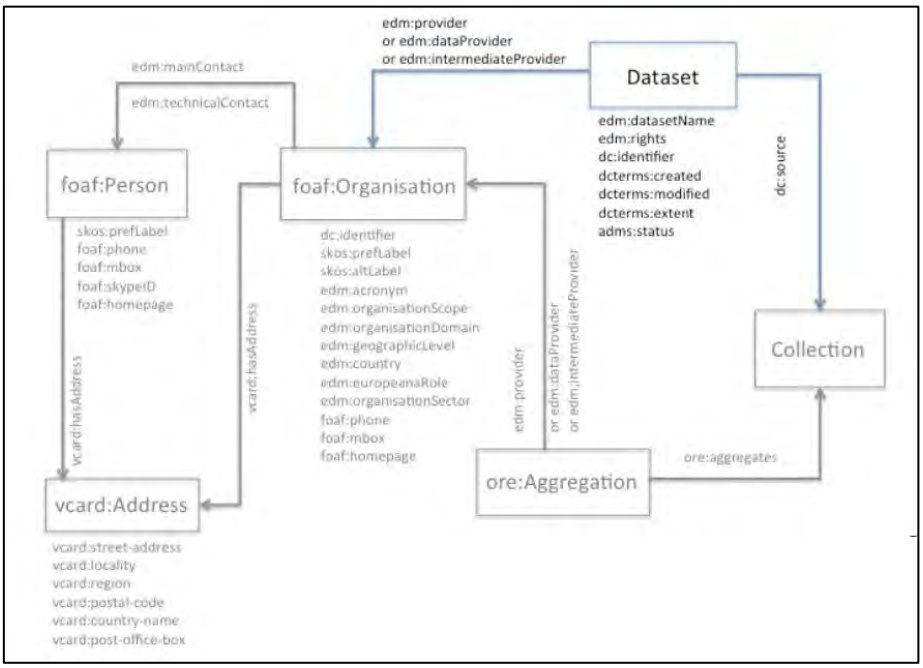


図 91 Europeana Dataset Profile の概要図

---

121 http://www.pref.akita.lg.jp/uploads/public/archive_0000008425_00/riyouannai.pdf (2017 年 1 月 24 日アクセス)

122 http://www.u-ryukyu.ac.jp/top_news/systemrenkei2013070901/ (2017 年 1 月 24 日アクセス)

123 http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_profiles/EDM_Dataset_Profile_042016.pdf (2017 年 1 月 24 日アクセス)

3.1.6. 利用サービス向上のための機能・サービス

デジタルアーカイブ・システムで提供する利用サービス向上のための機能・サービスの在り方を検討するため、全国の公文書館等における、デジタルアーカイブ・システムやホームページなど、インターネットを通じた機能・サービス提供の状況を調査した。

調査概要

調査目的	利用サービス向上のための機能・サービスの状況把握
調査対象	・国立公文書館デジタルアーカイブ ・アジア歴史資料センター ・国立公文書館のホームページにおいて、国内外や類縁機関等、国立公文書館に関係のあるウェブサイトを集めたリンク集に示されている、国内109 機関 ・国内の公文書館類縁機関
調査方法	・上記機関のホームページ等の利用(平成 28 年 12 月 19 日現在)
調査内容	・レファレンスサービスの状況 ・利用動向の解析の状況 ・その他特徴的な事例

(1) レファレンスサービス

本項では、オンラインまたはオフラインで利用者からの問い合わせに対応している事例について調査し、利用サービス向上に資する在り方を検討した。

(ア) 事例

尼崎市立地域研究史料館は国立国会図書館のレファレンス協同データベース[124]に参加している。利用情報を広く発信することを目的として平成 23 年 11 月に参加して以来、おおむね 2 週間に 1 件のペースでレファレンス事例を登録している。平成 27 年度、21 件の事例を登録した結果、平成 28 年 3 月末現在の登録件数はレファレンス事例 97 件、調べ方マニュアル 4 件となっている[125]。

尼崎市立地域研究史料館では、レファレンスを中心とする利用公開機能に力点を置き、利用者・市民とのインターフェースを重視する事業を展開している。1990 年代前半に行なった業務改革を起点とし、利用者の調査内容を聞き、閲覧調査の方法手順を説明し、必要に応じて史料検索を手伝う。単純なことだが、これを館内の共通事項とし、スタッフ全員で取り組んだ。その結果、利用者が大幅に増加した。下図のとおり、業務改革初期の 1990 年代前半と直近を比較すると、2 倍以上増加している。それ以前は利用者ゼロの日が珍しくなかったため、上記取り組みにより、利用が飛躍的に増大したと言える[126]。

---

124 図書館及び類縁機関におけるレファレンス情報の共有化を目的として、国立国会図書館が Web 上に運営する公開データベース

125 http://www.archives.city.amagasaki.hyogo.jp/about/annual/pdf/2016.pdf (2017 年 1 月 24 日アクセス)

126 http://203.180.136.64/about/publication/archives/pdf/acv_51_p26.pdf (2017 年 1 月 24 日アクセス)

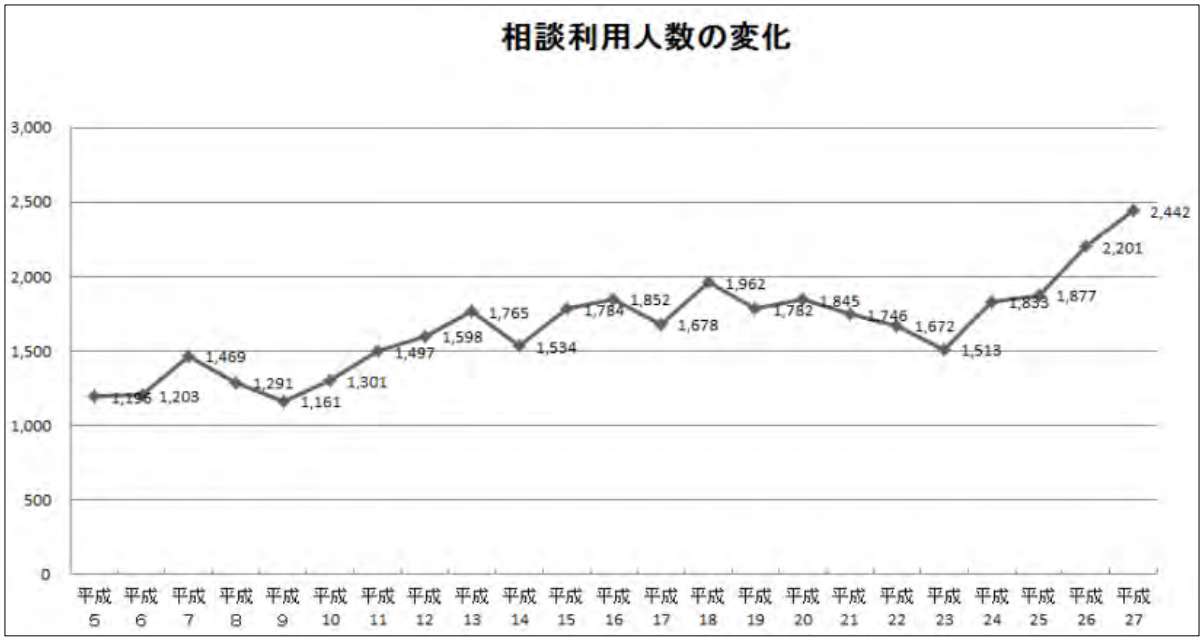


図 92 尼崎市立地域研究史料館の相談利用人数の変化[127]

国立国会図書館のレファレンス協同データベースには、尼崎市立地域研究史料館の他、外務省外交史料館、茨城県立歴史館（閲覧室）、東京都公文書館、福井県文書館、京都府立京都学・歴彩館、福岡市総合図書館、寒川文書館、芳賀町総合情報館が参加している。

（イ）考察

公文書館等における実態としては、来館者に対するカウンターでの対応が多いと思われる。しかしながら、年々増えていく自館の資料についてだけではなく、自館の資料に関連する他館の資料についての問合せがあった場合、自館内だけのデータの蓄積では回答できない。そのような場合に質問及び回答が蓄積されたレファレンス協同データベースのような仕組みが回答を手助けし、場合によっては問合せを受けたその場で他館にある旨を案内することも可能と思われる。

尼崎市立地域研究史料館の事例から、レファレンスを充実させることによって、来館者が増え、市民ボランティアとの協同作業のきっかけになっていくのではないかと思われる。レファレンス情報の提供など、サービスの実施を支援する機能をデジタルアーカイブ・システムが備えることは、利用サービス向上に有効であると考えられる。

また、アジア歴史資料センターのように利用者から目録データや誤字脱字といった不具合情報を受付ける仕組みを SNS も活用して拡大していくことにより、サービス向上は元より、利用者の満足度向上につながると考えられる。

(2) 利用動向の解析

本項では、デジタルアーカイブ・システムの機能・サービスの利用動向を解析する方法について調査し、

---

127 http://www.archives.city.amagasaki.hyogo.jp/about/annual/pdf/2016.pdf（2017 年 1 月 24 日アクセス）

サービス向上にどのように反映できるか検討する。

（ア）事例

- Google Analytics[128]

  1 か月あたりの上限ヒット数 1,000 万ヒットまで無料で使用できる(スタンダード)。Google Analytics では、ユーザーの行動や訪問経路を解析できる。

  Google Analytics のアクティビティストリームでは、共有された URL、共有の場所と方法（Google+で「再共有」、「投稿」、「コメント」など）、その URL に対するユーザーのコメントを確認できる。

- Yahoo! アクセス解析[129]

  Yahoo! プロモーション広告を利用している場合にのみ導入可能なサービスである。デジタルアーカイブ・システムへの導入には適さないと考えられる。

- Facebook ページインサイト[130]

  Facebook のアクセス状況解析サービスとして提供されている。30 人以上の「いいね！」が得られた場合に使用できる機能である。

- Twitter アナリティクス[131]

  Twitter のアクセス状況解析サービスであり、広告を利用している場合にのみ導入可能。デジタルアーカイブ・システムへの導入には適さないと考えられる。

これらのサービスを利用することで、コンバージョン(Web サイトの達成すべき目標)を定量的に示すことができる。

また、検索でよく利用されるキーワードを可視化することにより、利用者がどのような資料を求めているのか把握し前述のレファレンスの基礎情報とすることや、資料紹介・デジタル展示といったコンテンツを作成する際の指針とするなど、資料利用サービスの向上に結び付けることが考えられる。

（イ）考察

アンケート結果では、開示請求の対象とならない単純なアクセス件数等のログ情報で十分であるという回答や、全ての操作を記録するようなものではなく統計的に整理できるものでよいとの回答があった（3.1.2 参照）。しかし、利用者のサイト内での動向解析や分析を行うことによって、デジタルアーカイブ・システムの利用状況を把握し、利用サービス向上に向けた施策を見出せる可能性があることから、分析等結果を元に提供するサービスの改善等に活用できるものと考えられる。

---

128 https://www.google.com/intl/ja_jp/analytics/（2017 年 1 月 24 日アクセス）
129 https://analytics.yahoo.co.jp/（2017 年 1 月 24 日アクセス）
130 https://ja-jp.facebook.com/help/336893449723054/（2017 年 1 月 24 日アクセス）
131 https://business.twitter.com/ja/analytics.html（2017 年 1 月 24 日アクセス）

3.2. デジタルアーカイブ・システムへの適用に係る検討

本調査で調査したサービス等について、「導入の容易さ」「運用効率」「利活用促進効果」「課題」の観点により、デジタルアーカイブ・システムへの適用について検討する。デジタルアーカイブ・システムへの適用可能性、課題について検討した。

3.2.1. デジタルアーカイブの基本機能(利用者向け)

(1) 検索機能

(ア) キーワード検索、詳細検索

キーワード検索とはキーワードを指定して検索する機能、詳細検索とはキーワードの他に条件を指定して検索する機能を指す。

導入の容易さ	すべてのシステム導入機関で採用されており、オープンソースの検索エンジンやパッケージソフトウェアがあるため導入は容易である。
利活用促進効果	標準的かつ必須の機能であり、効果は低い。
課題	各機関の所蔵資料は形態面の多様性があり、目録項目や利用方法によって検索対象とする項目には差異がある。各項目で扱うデータの内容ごとに適切なデータ型を設定し、このデータ型に合わせて目録データを整備することも検索効率向上に向けた対応として考えることができる。

(イ) 階層検索

階層検索とは、公文書等に関する階層構成を選択しながら目録を検索する機能を指す。

導入の容易さ	階層検索はシステム導入機関の半数で何らかの形で採用されており、導入の難易度は低いと考えられる。
運用効率	組織改正が行われた場合に階層の付け替えや移行を行わなければならない可能性があり、運用効率に影響を及ぼす可能性がある。
利活用促進効果	対象となる資料が階層構造を持つことを考えると、資料の構造に詳しい研究者や行政経験者等利用者の検索の利便性を高め、一定の利活用促進の効果があると考えられる。
課題	階層を表すユーザーインターフェースはシステムごとに差異があり、利用者の利便性を損なう恐れがある。デジタルアーカイブとして押さえるべき仕様を明確にする必要がある。

（ウ）辞書検索

辞書検索とは、同義語や関連語等の辞書(シソーラス)を用いて、利用者が指定したキーワード以外の言葉により目録を検索する機能を指す。

導入の容易さ	導入済みの事例があり、辞書やシソーラスによる同義語検索機能を備えた検索エンジンやパッケージソフトウェアがあることから、導入の難易度は低いと考えられる。
運用効率	収集する資料の範囲が広がる等状況が変わるごとに、辞書データを整備する必要があることが、運用効率に影響を及ぼす可能性がある。
利活用促進効果	専門的な用語が使われる歴史的公文書では利用者の検索の利便性を高めることができるため、利活用促進効果は高いと考えられる。
課題	辞書として使用するデータは各機関の資料の内容に応じて整備が必要であり、そのためのコストが課題となる。

（エ）異体字・旧字検索

異体字・旧字検索とは、異体字や旧字体(漢字の「一」と「壱」、「国」と「國」等、ひらがなとカタカナ等)を同一のものとして検索する機能を指す。

導入の容易さ	日本語に対応した検索エンジンでは標準機能として用意されていることが多いが、設定については個々に検討が必要である。導入の難易度は検索エンジンに応じて変化しうる。
利活用促進効果	検索時のヒット率が向上し、利用者の検索の利便性を高めることができるため、利活用促進効果は高いと考えられる。
課題	辞書と同様に、異体字・旧字として扱うデータは各機関の資料の内容に応じて検討が必要である。一方で、公文書館等における共通定義を整備し、各機関が共有して利用する方法もシステムの運用面においては考えることができるものと思われる。

（オ）サジェスト機能

サジェスト機能とは、利用者が指定したキーワードの一部から、予測される検索キーワードの候補語を提示し、選択させる機能を指す。

導入の容易さ	技術的には検索エンジンの機能として実現されており、導入の難易度は低いと考えられるが、パッケージソフトウェア等の実装状況により異なる。
利活用促進効果	アンケート結果から、各機関でも利用者の検索の利便性を高める機能と期待されており、各機関の資料に初めて接する利用者にはどのようなキーワードで探せばよいかのヒントとなるため、利活用促進効果は高いと考えられる。
課題	利用頻度の高い専門的知識を持った利用者には逆に使いにくくなる可能性がある。本機能の適用の是非を検索時に利用者が指定できるように実現ことが望ましいものであると思われる。

（カ）ファセット機能

ファセット機能とは、検索結果を特定項目の内容ごとに分類して件数を表示し、利用者が容易に絞込みを行える機能を指す。

導入の容易さ	技術的には検索エンジンの機能として実現されており、導入の難易度は低いと考えられるが、パッケージソフトウェア等の実装状況により異なる。
利活用促進効果	海外の事例調査では実装率は高く、利用者の利便性向上につながることから一定の効果が期待できる。
課題	ファセットを効果的に使うには、対象となる項目の選定とデータの整理が必要となる。

（キ）パーマリンク機能

パーマリンク機能とは、個々の目録情報やデジタルコンテンツに対して永続的な固有の URI を与えることにより、利用者が該当の情報に容易にアクセスすることができる機能を指す。

導入の容易さ	永続性を考慮した識別子と URI の設計を行う必要があり、導入の難易度は高い。
利活用促進効果	本調査検討で取り上げた RDF や LOD、IIIF といった技術において根幹となる要素であり、相互運用性や機械可読性の向上による利活用促進効果が期待できる。
課題	目録やデジタルコンテンツの識別子は、各機関の過去のデータ整備の中で付与されてきたものであり、パーマリンクとして新たな情報を付与し、管理する必要があるケースが想定される。 また、その一貫性や恒久性を保証するためには、各機関でシステム更新を行う際に URI が維持されることを明示的に仕様として入れる必要がある。

(2)　横断検索機能

横断検索機能とは、自館のデータベースと外部機関のデータベースを一度の操作でまとめて検索し、結果を取得できる機能を指す。

導入の容易さ	検索機能を提供するパッケージソフトウェアで、横断検索において標準的に用いられているプロトコルをサポートしている場合、導入は容易である。
利活用促進効果	複数の機関に跨って関連する資料が分散されている場合があり、それらを横断的に検索できることは利用者の利便性向上につながり、利活用促進が期待できる。
課題	上記プロトコルに対応していないパッケージを用いる場合、実装のためのコストがかかる。 横断検索で用いる目録項目の標準化が必要となるが、利用者の検索精度と提供する情報の充実度の向上のためには、より多くの目録項目に対応する必要がある。

(3)　画像配信・閲覧機能

（ア）画像配信機能

画像配信機能とは、資料画像(デジタルコンテンツ)を利用者へ配信する仕組みを指す。

導入の容易さ	PDF や JPEG といった一般的な形式での配信以外に、高精細画像を配信するためのオープンソースソフトウェアやパッケージソフトウェアもあり、導入は容易である。 「2. デジタルアーカイブ・システムに関連する最新 IT 技術に係る調査検討」で触れた通り、デジタルアーカイブでの画像公開・共有のための国際的な規格として IIIF の採用が広まっており、利用者の利便性と相互運用性の向上のために対応した仕組みを導入することが望ましい。
利活用促進効果	各機関での原本や複製物の閲覧以外に、資料のデジタルデータをインターネット上で利用できることは地理的に離れた利用者の利便性向上につながり、利活用促進が期待できる。
課題	技術的なハードルは低いがパッケージソフトウェアやサービスが対応していない場合は配信機能の実装にコストがかかる。各機関におかれては、利用者に提供する画像の種類や画質、サイズを踏まえて、対応を検討することになるものと考えられる。

（イ）画像閲覧機能

画像閲覧機能とは、配信されたデジタルコンテンツを利用者が閲覧するために使用する機能を指す。

導入の容易さ	HTML5 や JavaScript といった技術を組み合わせた、高機能な画像閲覧ソフトウェア（ビューア）のライブラリがオープンソースで公開されており、ビューアの実装にあたっての技術的なハードルは低く、導入は容易である。
運用効率	IIIF 等システムへの依存度を軽減させることが期待できるアプリケーションソフトウェアの採用により、利用者に対する画像閲覧サービス提供に係る運用において、運用効率の向上を図ることが可能と思われる。
利活用促進効果	拡大・縮小や移動、印刷といった機能を持つ高機能ビューアを導入することで利活用促進の効果が期待できる。
課題	技術的なハードルは低いがパッケージソフトウェアやサービスが対応していない場合はビューアの実装にコストがかかる。画像配信と合わせた検討と選択が必要であると考える。

（ウ） 動画・音声配信機能

動画・音声配信機能とは、動画や音声等のデジタルコンテンツを利用者へ配信する仕組みを指す。

導入の容易さ	ストリーミング配信を行う場合は専用のサーバソフトウェアの導入が必要となり、技術的・コスト的なハードルは高く、導入の難易度は高い。
運用効率	配信対象となるデータ量によっては運用効率に影響を及ぼす可能性がある。
利活用促進効果	アンケート結果より、動画・音声の再生機能には、各機関における、一定のニーズがあることを確認。利活用促進に一定の効果が期待できるものと考える。
課題	配信するデータ量によっては、上記の技術的・コスト的な課題がある。公開する動画・音声（配信するデータ量）が少ない場合には、外部の事業者が提供する動画配信サービスを利用することでこれらの課題を解消できる。しかし、同サービスの利用においては、機関の外部へコンテンツを配置することになるため、セキュリティポリシーを踏まえた検討が必要であると考える。

（エ） 画像の読み下し文表示機能（複数資料の重ね合わせ）

古文書等に書かれた文字を、現代語に直した読み下しテキストとして画像に重ねあわせて表示する機能を指す。

導入の容易さ	テキストを重ねあわせて表示するビューアが必要となり、読み下しテキスト作成も含めて導入の難易度は高い。
運用効率	画像ごとに読み下しテキストの作成が必要となるため、業務負担が増えるとともに、システムの運用効率に少なからず影響を及ぼすものとなると考えられる。
利活用促進効果	専門知識を要する古文書の解読が容易になるという点では、利用者の裾野を広げることになるため、一定の利活用促進効果が期待できる。
課題	技術的なハードルもあるが、読み下しテキストを作成すること自体が高い専門知識を有する作業である。

3.2.2. デジタルアーカイブの基本機能(業務)

(1) 目録データの管理・登録機能

(ア) データの登録、編集、削除機能

導入の容易さ	パッケージソフトウェアやオープンソースの RDBMS 等を組み合わせることで実現可能であり、導入は容易である。
運用効率	実現される機能により運用効率には幅があると考えられる。職員の作業負担を軽減するためには、汎用的なフォーマットで作成された目録データにより容易にシステムへの登録・更新が行えるユーザーインターフェースと機能が必要である。
課題	目録データは各機関の方針によりデータの意味や処理方法に違いが生じうるものであり、機能への要求レベルも異なるため、汎用化が難しく、各機関独自のカスタマイズや機能追加によりシステム構築にかかるコストが増加する可能性がある。目録項目の標準化と、機能の内容を取捨選択することで、機関の人員や予算に応じたデジタルアーカイブの構築が可能になると考えられる。

(イ) 公開・非公開機能

公開・非公開機能とは、目録データを一般利用者へ公開するか、非公開とするかを制御する機能を指す。

導入の容易さ	パッケージソフトウェアやオープンソースの RDBMS 等を組み合わせることで実現可能であり、導入は容易である。
運用効率	本機能を実現する際の画面表示等(ユーザインタフェース)の設計により、運用効率への影響には幅が生じるものとなると考えられる。 非公開の目録を職員のみが検索・閲覧できる機能は多くの機関が求めていることから必要な機能といえる。
課題	非公開情報を扱うことから、サーバ、データベースやアプリケーション等システムの各構成要素における充分なセキュリティの確保が課題となると考えられる。 また、一般の利用者のみではなく、各機関の職員による利用においても、担当職務等に応じた目録データの閲覧制限を要する場合が考えられる。この場合は、別途運用方法を考慮する等対策を講じる必要があるものと考えられる。

(2)　画像データの管理・登録機能

（ア）画像の管理・登録機能

導入の容易さ	パッケージソフトウェアで実現されている機能であり、技術的なハードルは低く、導入は容易である。 ただし、動画・音声配信機能でも触れたとおり、動画・音声といったデジタルコンテンツを扱う場合は、技術的・コスト的なハードルが高くなることから、導入の難易度が高くなる。
運用効率	実現される機能により運用効率には幅があると考えられる。
課題	各機関の実情に見合った機能内容を検討する必要があるものと考える。

（イ）公開・非公開機能

公開・非公開機能とは、画像等のデジタルコンテンツを一般利用者へ公開するか、非公開とするかを制御する機能を指す。

導入の容易さ	アプリケーションとしてはデータベース及びファイルの制御で実現できるため、技術的なハードルは低く、導入は容易である。
運用効率	実現される機能により運用効率には幅があると考えられる。
課題	目録データ同様に、サーバ、データ格納領域、アプリケーション等システムの各構成要素における充分なセキュリティの確保が課題となると考えられる。 また、一般の利用者のみではなく、各機関の職員による利用においても、担当職務等に応じた資料画像等の閲覧制限を要する場合が考えられる。この場合は、別途運用方法を考慮する等対策を講じる必要があるものと考えられる。

(3)　利用状況把握機能

利用状況把握機能とは、利用者によるシステムやシステムが提供する各コンテンツ等へのアクセス状況に関する情報を収集し、分析等処理に供することにより、利用者の利用動向等状況を把握するための仕組みである。

導入の容易さ	オープンソースの Web サーバログ解析ツールや(条件はあるが)無償で利用できるサービスが民間の事業者により提供されており、導入は容易である。
運用効率	取得する情報により、運用効率の良し悪しに幅が生じるものであると考えられる。 また、利用状況についての定量的な指標値を得ることにより、アクセス数の増加に対応するためのリソース増強や機能改善に向けた計画を立案する助けになる。
課題	民間の事業者が提供する、外部のサービスを利用する場合は、期間内の情報を外部へ提供することになりうるため、各機関のセキュリティポリシーの制約を受ける可能性がある。

(4)　閲覧業務関連機能

閲覧業務関連機能のうち、利用者に対するサービスに関する主要な機能として、利用請求管理機能について検討する。利用請求管理機能とは、利用者が検索システム上で利用請求書や複写申込書を出力した際に、その情報を職員が管理するための仕組みである。

導入の容易さ	データベース機能として実現できるものであり、技術的なハードルは低いが、閲覧に関する業務は各機関の運用方針により異なるため、導入の容易さには幅があるものと考えられる。
運用効率	閲覧や複写等の利用請求を、目録データベースと連携させることで職員の作業負担軽減につながり、システムの運用効率向上が期待できるものであると考える。ただし、導入する機能の内容により、運用効率の良し悪しには、幅が生じるものとなると考えられる。
課題	閲覧に関する業務は公文書館等に共通の業務であるが、各機関に求められる役割や運用方針により、システム化にあたって独自開発を要する等コスト上昇に結びつく可能性がある。閲覧業務に共通のエッセンスを抽出し、標準となるひな形や機能を定めることにより、コストの低減を図ることができるものであると考えられる。

3.2.3. 利用者拡大のための機能・サービス

(1) 利用者による情報発信支援

(ア) SNS 連携

SNS 連携とは、利用者がデジタルアーカイブ上の目録やデジタルコンテンツの情報(URL)を不特定多数の利用者へ発信することを支援する機能を指す。

導入の容易さ	主要な SNS では連携のための Web API が公開されており、技術的なハードルは低く、導入は容易である。
利活用促進効果	デジタルアーカイブの利用者増加につながり、利活用促進が期待できる。
課題	SNS サービスの流動性は高く、現在主流となっているサービスであっても、数年で利用者が減少する可能性や、サービス停止となる可能性があるほか、後発の新しいサービスが主流となることも想定される。 また、各機関のセキュリティポリシーの制約を受ける可能性や、利用者同士のトラブルに各機関が巻き込まれる恐れがあることも想定可能である。

(イ) ソーシャルタギング

ソーシャルタギングとは、システムが提供するコンテンツ等の情報に対し、一般の利用者がレッテルを貼る(タグ付けする、またはメタデータを登録する)ことを支援する機能を指す。

導入の容易さ	SNS 等ソーシャルメディアでは広く採用されている機能であり、技術的なハードルは低く、導入は容易である。
利活用促進効果	利用者がシステムの提供するコンテンツ等に対するメタデータを能動的に登録することを支援する機能であり、システムが提供する情報の整備へ積極的に関わる利用者の出現につながるものと考えられることから、利用者にとってより身近なシステムとなりえることにより、システムの利活用促進が期待できるものと考えられる。
課題	ソーシャルタギングは利用者が内容を理解し、適切なタグを付与することで有効な情報検索手段ともなりうるが、タグとして用いられるキーワードは統制されず、利用者ごとに異なることが懸念されるため、他の利用者を誤った結果へ導いたり、混乱を生じさせたりする恐れがある。 上記において、各機関が提供する情報を補助するものとして有用なタグが適切に登録されるようにするための対応の検討を要する場合もあるものと考えられる。

(2)　利用者によるシステムへの能動的参加

導入の容易さ	技術的な難易度は高くないが、利用者による業務支援の観点から、運用上の制約について考慮する必要があり、導入の難易度は高い。
運用効率	職員がデータ作成にかける時間と手間は減るが、投稿されたデータの査閲・承認にかける時間と手間が増えるため、サービスの運用面における効率に影響を及ぼす可能性がある。
利活用促進効果	対象となる利用者による宣伝効果が期待できる。
課題	職員(アーキビスト)の登録したデータと明確に分ける必要がある。投稿されたデータの査閲・承認、投稿者がシステムを利用する場合のアカウント管理やセキュリティの確保等、技術的課題と、品質確保の手段、責任の所在等、運用にあたっての課題がある。デジタルアーカイブ・システムへのデータ登録の面では、上記課題を鑑み、実現可能性について具体的な検討が必要と思われる。

(3)　デジタルコンテンツの活用

（ア）デジタル展示による活用

デジタル展示とは、デジタルアーカイブ・システムが提供する資料を、一定のテーマにより集積し、目録以外の解説情報を付加したうえで公開する仕組みを指す。

導入の容易さ	デジタルアーカイブ・システムが提供する目録データや画像データの URL を、デジタル展示として提供する情報（コンテンツ）に埋め込むことで、システムとデジタル展示との連携は容易であり、技術的ハードルは低い。
運用効率	コンテンツを用意するためには、資料に対する専門的知識とコンテンツ作成に対する専門的知識を要するほか、コンテンツの作成費用や期間への考慮も要することから、システムの運用に影響を及ぼす可能性がある。
利活用促進効果	利用者が興味を持つテーマを選ぶことで目録やデジタルコンテンツに誘導することができ、利用の促進が期待できる。
課題	目録データや画像データによる連携には、システム更新時等におけるリンク切れの可能性を考慮すべきと考える。この点を回避するためには、これらのデータに対する URL として、パーマリンクを用いる方法が考えられる。

（イ）SNS によるコンテンツ活用

SNS によるコンテンツ活用とは、SNS の公式アカウントで、デジタルアーカイブ・システムが提供する資料を利用者に発信して紹介する仕組みを指す。

導入の容易さ	SNS のアカウントを作るのみであり、導入は容易である。
運用効率	情報発信の頻度によっては、作業負担の増加からサービスの運用、業務効率に影響を及ぼす可能性があるものと考える。
利活用促進効果	SNS を通じて不特定多数の利用者へ情報を発信できることから、利用の裾野を広げ、利活用促進が期待できる。
課題	各機関のセキュリティポリシーの制約を受ける可能性があるほか、発信する内容によっては、予想しない非難を受ける（いわゆる「炎上」）状態になる恐れがあるこ

	とも想定される。

(4)　他機関との連携

（ア）類縁機関との連携

導入の容易さ	連携対象機関との調整が必要となり、導入の難易度は、連携内容に応じて変化しうる。
運用効率	連携内容によっては、連携を実現する特定の者に連携作業の負担が偏在することになる場合もあると考えられる。この場合、負担の少ない者にとってはシステムの運用効率が高まるものの、負担の大きな者にとっては運用効率が低くなることも考えられる。
利活用促進効果	一つの機関が所蔵する資料数を上回る数の資料について、場合によっては、付加情報を付けての利用が可能となることから、利用者に対する利便性が向上し、利活用の促進が期待できる。
課題	事前の調整に時間とコストがかかる場合がある。また、連携方法によっては技術的な課題が発生する場合も考えられる。

3.2.4. 利用サービス向上のための機能・サービス

(1) レファレンスサービス

レファレンスサービスに関する機能のデジタルアーカイブ・システムへの導入について、以下に検討する。

導入の容易さ	レファレンスサービスに関する機能として、レファレンスに係る問い合わせの受付、蓄積、管理、回答支援等多様な機能が想定可能である。各機能の実現において、技術的なハードルは内容により変わるものであり、コスト面も含め、導入の容易さには幅があるものと考えられる。
運用効率	レファレンスに関する情報を蓄積、共有することにより、レファレンスサービスの運用効率の向上が期待できる。
利活用促進効果	利用者の問い合わせに対する回答時間の短縮による利便性の向上から、一定の利用サービスの向上が期待できる。
課題	本サービス運用におけるシステム化は、関係する多様な業務でのワークフローの整理を伴うものであり、整理・調整に関する作業負担が課題となるものと考えられる。

(2) 利用動向解析機能(システムのログ解析)

利用動向の解析機能について、利用状況の把握に加えて、様々な解析手法を用いて利用者の動向を分析する仕組みのデジタルアーカイブ・システムへの導入について以下に検討する。

導入の容易さ	民間の事業者が提供する、無償で利用できるサービスがあり、導入の難易度は低い。
利活用促進効果	利用者動向から将来のアクセス数増加につながる施策を打ち出すことにより、利用サービスの向上が期待できる。
課題	分析結果の解釈に知識や時間が必要となる可能性があるものと考えられる。 また、外部の民間事業者が提供するサービスの利用は、各機関から外部への情報提供を伴う可能性があり、各機関におけるセキュリティポリシーの制約を受ける可能性があるものと考えられる。

4. まとめ

4.1. 検討結果

4.1.1. 「デジタルアーカイブ・システムに関連する最新 IT 技術に係る調査検討」の調査検討結果のまとめ

(1) システム構成等の調査検討結果

デジタルアーカイブ・システムのシステム構成は、システム基盤の類型（オンプレミス型、コロケーション型、プライベートクラウド、パブリッククラウド）を基本として、アプリケーションの提供方法や運用形態を考慮して考えることができる。運用形態では、従来から行われてきた一つの機関が単独で構築・運用する形態（単館運用）の他に、複数の機関が共同して構築・運用する形態（共同運用）が考えられる。

単館運用の場合、業務や利用者の要求にあわせたアプリケーションを構築し、サービス面において機関の独自性を出すことが比較的容易であるが、予算や人員を独自に確保する必要がある。また、システムの設置形態として、一般にはオンプレミス型、コロケーション型（データセンター活用型）が考えられるが、デジタルアーカイブ・システムが、インターネットを通じて 24 時間サービス提供することを前提とすると、安定した電源や回線等を備え、運用担当者が 24 時間体制で監視をするデータセンターを利用することがより望ましく、コロケーション型がより適していると言える。

共同運用の場合、利用者の利便性向上という点からは、単にシステム基盤が共有されているというだけではなく、アプリケーションにおいても共通した機能の利用が求められる。アプリケーションの開発・導入やパッケージ導入においては、利便性向上につながる、共同運用ならではの機能を織り込むことや、共同運用する機関の事情に配慮するなど、柔軟な対応を要する。このような付加価値を提供する観点からは、システム基盤としてはパブリッククラウドよりも、専用のクラウド環境であるプライベートクラウド（IaaS、PaaS）の採用が適していると考えられる。一方、費用や人員等の負担軽減という点からは、サービス利用型（SaaS）システムの活用が有効な選択肢である。サービス利用型のデジタルアーカイブ・システムを利用する場合は、パブリッククラウドを採用する方が、提供されているサービスの選択肢も多く、コストメリットが現れやすい。

(2) クラウド等 IT 技術の調査検討結果

クラウド技術については、低コストでシステムを導入できることから有用であり、様々なサービスが商用ベースで提供されている。しかし、データのセキュリティや活用可能なアプリケーション機能等の運用環境において、それぞれの機関が提供したいと考えるサービスの実現可能性や運用方法等の十分な検討を要するものと考えられる。

サーバ環境では、サーバ技術とストレージ技術について調査をした。サーバ機器、OS、ミドルウェアは、低価格化とオープンソース化が進んでいる。また、サーバ機器の性能向上に合わせて仮想化技術も発展している。ストレージの容量あたりの単価は減少傾向にあり、大容量のデータを扱うためのコスト面でのハードルは低くなりつつある。また、クラウドストレージを使うことで、提供されるサービス内容に制限はあるものの、必要な容量のみをコストを抑えて用意することが可能である。このような状況から、デジタルアーカイブ・システムを実現するためのサーバ環境は年々高機能かつ低コスト化が進んできていることが伺え、サーバ環境（ハードウェア面）の整備におけるコスト面でのハードルは低減しているものと考えられ

る。

ネットワーク環境では、インターネット接続回線と通信プロトコルについて調査した。インターネット接続回線は、公開するコンテンツ、想定のアクセス数、予算に応じた適切な回線を用意することが必要である。ただし、各機関がすでにインターネット回線を使用しており、その回線が相当の要件を満たしている場合は、セキュリティに配慮して、既存の回線の活用を図ることが合理的な対応であると考える。

通信プロトコルにおいては、HTTP の最新版であるHTTP/2 への対応により、閲覧速度向上により利用者のストレス軽減という効果が期待できる。また、利用者やサービス提供者（各機関）が送受する情報に係るセキュリティ等を考慮すると、暗号化した通信を提供する HTTPS への対応が望ましいと考えられる。

バックアップ関連技術には、複数のバックアップ方式、バックアップメディア等がある。各機関の保有するデータ量と増加予測、データの特徴に応じて、適切なバックアップ方式やメディアを検討し導入することが推奨される。

ネットワークセキュリティ技術では、コンピュータネットワークにおける脅威に対抗する様々な技術やサービスが提供されている。しかし、これらの技術等の導入に係るコストは、セキュリティレベルに応じて高くなる傾向にある。各機関のセキュリティ対策方針に応じて、IPA 等の公表資料で示される対策を参考に適切な対応を選択する必要がある。

目録関連技術では、符号化記録史料記述の国際的な標準の最新版であるEAD3 への対応が望ましい。横断検索の項目には、アーカイブズ機関以外との横断検索を視野に入れると、Dublin Core を採用することは有用である。Dublin Core の基本的な 15 エレメントは簡潔であることから多様な領域に適用できるが、表現可能な情報の粒度が粗く、メタデータの相互運用性を低下させる恐れがある。そのような不足を補うものとして、Dublin Core の拡張項目であるDC タームスでは、より精度の高い情報の記述が可能となっている。横断検索を実現する機能の運用や検索精度の面から、最低限共通する目録データ項目を設定した上で、連携する目録データ項目を整理することやアプリケーションプロファイルの考えを取り入れることで相互運用性を高めることが重要である。また、デジタルアーカイブ・システムにおいて、システム移行等を経ても、目録データ等の提供情報を長期的に安定して利用可能とするパーマリンク（固定 URL）の導入は、同システムにおける目録情報等を識別する技術であるとともに、利用者の利便性を損ねないための技術として有用である。その他目録関連の標準化の動向として、国際公文書館会議で検討されている新しい記述標準である RiC(Records in Contexts)の動向も注視すべきである。

目録データの提供では、人間が利用するためのデータ提供と、機械可読性の高い形式によるデータ提供の両者に対応することが推奨される。後者においては、デジタルアーカイブ・システムが提供するデータを別のアプリケーションプログラムで取得、活用可能とすることにより、利用者や他機関が作成するアプリケーションから、各機関の目録データやデジタルコンテンツデータを利用することが可能となり、利活用の拡大を期待することができる。更にデジタルアーカイブ・システムで提供するデータは、今後はオープンデータへの対応が求められると考えられ、オープンデータを実現する RDF や LOD といった技術への対応や、アーカイブズ所蔵機関に固有の識別子を定め、維持することなどが望まれる。また、利用者が安心してデジタルアーカイブ・システムが提供するデータを利用できるようにするためには、政府標準利用規約やクリエイティブ・コモンズといったライセンスに関する標準への準拠と提供データにおける表示に対応することが望ましい。

情報検索技術に関しては、デジタルアーカイブ・システムで扱うデータは非構造化データ等多様化が進んでおり、このような状況における情報検索の手段として、全文検索への対応がある。全文検索に対応する仕組み（全文検索エンジン）には、高機能・高性能なものがオープンソースでも提供されており、導

入のハードルは決して高いものではなくなってきている。また、全文検索機能を有する RDBMS を用いることにより、データ管理の効率化を図る方法も考えられる。

横断検索では、SRU/W、OpenSearch のいずれかを実装することが、相互運用性の向上という点からも推奨される。またオープンデータ推進の観点では、機械可読形式でのデータ提供と併せて Web API を提供することが推奨される。

デジタルコンテンツの配信では、デジタルアーカイブで扱うコンテンツの多様化を踏まえて、動画・音声等マルチメディアの配信について検討する必要がある。これらの配信方式としては、ストリーミング方式やダウンロード方式がある。動画・音声の配信サービスでは、データ保護等の観点から、ストリーミング方式によることが一般的であるが、配信に専用のソフトウェアを必要とする。また、配信サーバの負荷軽減や利用者の閲覧等におけるレスポンスの向上を期待することができるコンテンツキャッシュ技術の適用が可能である。しかし、配信ソフトウェアやコンテンツキャッシュ技術の導入は、導入コストや運用コストの発生を伴うことから、費用対効果を踏まえた検討が必要であると考える。また、国際的な画像配信のためのフレームワークであるIIIFは、デジタルアーカイブ・システム間の相互運用性が高まるものとして期待されている。

クライアント環境では、普及の著しいスマートフォン、タブレットといったスマートデバイスによる利用の増加が予想されるため、マルチデバイスに対応したユーザーインターフェースを備えることが望ましい。近年主要なブラウザは HTML5 や CSS3/4 に対応しており、ブラウザのみで高機能なユーザーインターフェースを実現できることに加え、多様なデバイスに対応したユーザーインターフェースを実現する手段のひとつである、レスポンシブウェブデザインを採用することも可能である。また、高速モバイル通信を用いたスマートデバイスの利用が増加しており、クライアント環境として今後も考慮すべきところであるが、通信回線におけるトラフィックが増大していることに対し、月当たりのデータ通信量の上限を設ける等の対応を実施する通信キャリアが出てきている。利用者の利便性を損なわないように配慮した対策等を講じられることが望まれる状況にある。

(3) デジタルアーカイブ・システムに活用可能と考えうる技術の調査検討結果

OCR 技術と画像検索技術については、高い解読正解率を取得可能な文書は、公文書館が所蔵する資料からすると一部の文書に止まるものの、OCR の文字読取り精度は向上している。公文書館の所蔵資料への OCR 技術の適用可能性については、処理対象資料に対する認識試験を重ね、判断することになると考える。また、OCR 技術と画像検索技術と組み合わせたくずし字解読等の実用化には、まだ時間がかかると推測される。今後の研究の進展やかかる技術の発展を注視し、導入可能な場合に追加すべきと考える。

自動翻訳については、日本語から複数の言語へ、また複数の言語から日本語への自動翻訳技術の研究は進められている。ただし、公文書や歴史的資料には特有の表現があり、対訳辞書等の情報整備が必要と考えられる。

辞書やシソーラス機能と利用可能な辞書データについては、デジタルアーカイブ・システムでも辞書検索を備えた事例はあるが、公文書や古書・古文書といった資料を効果的に検索できる辞書データは一部の機関でしか整備されていない。これらの資料の特徴を踏まえた辞書やシソーラスの整備にかかるコストを抑えるためにも、各機関の持つ辞書データを共通化し、相互利用できる仕組みが期待される。

4.1.2. 「デジタルアーカイブ・システムで提供するサービスに係る調査検討」の調査検討結果のまとめ

(1) 全国の公文書館等へのアンケート結果

全国の公文書館等へのアンケートから、全国の公文書館等、とりわけ市区町村レベルではシステム化の予算と人員の確保が最大の課題となっていることを確認した。また、機関が保有するデータを外部で管理することに制度上の制約があるとした機関がある一方、インフラとアプリケーションを共有してもよいと回答した機関が 6 割以上であり、その課題の解決策として、アプリケーションも含めた安価な共同利用型システムの活用が期待されている機関も少なからず確認された。ただし、運用形態で単館運用を希望する機関も 18%と一定数あり、機関の要望に応じて適切なシステム構成や運用形態を選択できることが必要であることが確認された。

整備が期待される機能（サービス）のうち、利用者向けの機能としては、検索キーワードのサジェスト機能や、異体字等の検索機能など、検索支援機能、動画・音声資料の提供機能へのニーズがあった。

業務面では利用請求記録の管理機能のほか、レファレンス業務を管理する機能や、職員による非公開データの検索・閲覧（資料の利用制限機能）への要望が大きいため、共同利用型システムで実現する機能を検討する際にはこれらが重要な要素であると言える。

(2) 国内外のデジタルアーカイブ・システムの調査結果

目録データの検索サービスをインターネット上で提供している機関は調査対象の 5 割弱であり、デジタルアーカイブといえど決して当たり前に提供されているサービスではないことを確認した。一方、5 割を超える機関で所蔵資料のデジタル画像を公開しており、利用者にアピールするコンテンツとして、広く所蔵資料のデジタル画像が捉えられている状況であることを確認した。

デジタルアーカイブの基本機能(利用者向け)では、基本的な検索機能に加えて、辞書機能、サジェスト機能、ファセットによる絞り込み検索機能といった事例があり、多様な検索支援手段が提供されていた。また、複数資料の重ね合わせやレスポンシブウェブデザインの採用など、ユーザーインターフェースの高機能化が見られた。

デジタルアーカイブの基本機能(業務)では、目録データ登録管理機能、画像データ管理機能、利用状況把握機能、閲覧業務関連機能の状況を調査し、必要な機能を把握した。

利用者拡大のための機能・サービスでは、SNS 等ソーシャルメディアは一部機関で活用していることを確認した。また、利用者からの提供情報（メタデータ・画像データ）を直接的・間接的に登録する事例としては、市民ボランティアによる目録データの整備事業や登録データの誤り等システム上の不具合を通報する機能の実装例を確認した。その他、デジタルコンテンツの展示・広報への活用事例、他機関との連携の事例を確認した。

利用サービス向上のための機能・サービスについては、レファレンスサービスの充実が利用者の増加につながる事例や、利用動向（システムログ）の解析サービスの事例を確認した。

(3) アンケート及び調査結果の検討

アンケートへの回答で示されたサービスや調査したサービスについて、それぞれ項目別に分け、デジタルアーカイブ・システムの適用可能性について検討した。以下に項目別の検討状況を示す。

デジタルアーカイブの基本機能(利用者向け)では、基本的な検索機能に加えて、辞書検索、サジェスト

機能、ファセット機能といった検索支援手段の整備により、利用者の利便性向上を図ることが重要である。ただし、これらの検索支援機能の実装においては、想定する利用者に合わせた内容とするよう検討が必要である。また、マルチメディアへの対応や、利用環境の充実を図る機能を整備することが必要である。

デジタルアーカイブの基本機能(業務)では、汎用的なフォーマットで作成された目録データにより容易にシステムへの登録・更新が行えるユーザーインターフェースと機能、目録データを個別に登録・編集できる機能、画像データを容易にシステムへ登録・更新・削除が行えるユーザーインターフェースと機能が必要である。また、目録データやデジタルコンテンツに対する利用制限を行う機能、利用請求管理へのニーズが高いことが確認された。各種業務に即した業務システムの整備が求められている。ただし、各機関のデータ項目や機能への要求レベルには差異があるため、目録項目等の標準化や、機能の内容の取捨選択等、各機関における業務の見直しから着手することで、機関の人員や予算に応じたデジタルアーカイブの構築が可能になると考えられる。

利用者拡大のための機能・サービスでは、利用者による情報発信に寄与する機能が有効である。利用者が SNS を用いて情報発信を容易に行える仕組み、各機関が SNS を用いて資料の情報を利用者に発信する仕組みが効果的である。なお SNS の活用にあたっては、組織の SNS 利用のガイドラインに準じる必要がある。また、デジタルアーカイブ・システムが提供する情報の充実を目的とした、利用者によるメタデータ付与等への能動的参加は、十分検討に値すると考えられる。データの品質や責任を明確にしたうえで、近年広まりつつあるクラウドソーシングを活用することも低コストでデータ化を行う選択肢としては有効と考えられる。この機能を実現するためには、利用者の作業環境実現における技術的な対応への考慮と、利用者による作業の管理等運用における考慮(運用方針の明確化)を要すると考えられる。その他、資料情報の展示等への二次利用は利用者拡大に有効と考えられる。

利用サービス向上のための機能・サービスでは、レファレンスサービスの充実が利用者数の増加につながる事例が確認されており、レファレンスサービスの実施を支援するデータベース構築等が有用である。単館運用よりも共同運用に適しているが、共同運用型システムでの運用にあたり、運用主体や費用分担が課題と考えられる。また、利用サービス向上の施策を検討するにあたり、デジタルアーカイブ・システムのログの解析による利用動向の把握は有効と考えられる。

4.2. 標準仕様書の改訂に向けて

国立公文書館では、全国の公文書館等におけるデジタルアーカイブ化の推進に資するため、平成 21 年 3 月に「全国の公文書館等におけるデジタルアーカイブ・システムの標準仕様書」（以下「標準仕様書」という。）を作成し、全国の公文書館等へ配布するとともに、訪問説明等を実施してきた。その後、平成 24 年 10 月に一部改訂しているが、標準仕様書の前回の改訂から 4 年以上が経過しており、デジタルアーカイブ・システムを取り巻く環境は大きく変化している。

本項は、「4.1.検討結果」において明らかとなった「デジタルアーカイブ・システムに関連する最新 IT 技術」および「デジタルアーカイブ・システムで提供するサービス」の調査検討結果を踏まえて、標準仕様書の改訂に向けた考え方を取りまとめるものである。

なお、改訂に向けた考え方を取りまとめるにあたっては、「国立公文書館の機能・施設の在り方に関する基本構想」(平成 28 年 3 月 31 日)等において、クラウド技術等を活用した共同利用型システム等の導入によるネットワーク化の推進が期待されている状況等を鑑み、以下の点に留意するものとする。

- クラウド技術の進展を踏まえ、システム基盤にクラウド環境を用いた構成の追加を検討する。
- 単館運用に加えて、共同運用を想定する。
- 国内外の規格等の動向を踏まえ、国内外の標準規格等への準拠を検討する。

4.2.1. 標準仕様書の改訂に向けた考え方

今回の調査結果を踏まえ、デジタルアーカイブ・システムの標準仕様書の改訂に向けた見直しのポイントは以下の通りである。

表 22 標準仕様書の改訂に向けた見直しポイント

デジタルアーカイブ・システムの標準仕様項目	見直しポイント
(1) システム環境	・各機関の要望に応じたシステム構成や運用形態を選択可能とする。 ・クラウド環境を活用する際の留意点、考えられるセキュリティ対策を示す。
(2) 目録情報のデータ化	・国内外の標準規格等の最新版への準拠を考慮する。
(3) 目録情報のデータベース化	・様々な検索支援機能を選択可能とする。 ・各種業務をシステム化する際の考え方を示す。
(4) デジタルコンテンツの取扱い	・音声や動画の配信を考慮する。
(5) インターネットにおける情報提供	・目録やデジタルコンテンツへのアクセスの永続性を考慮する。 ・多様な利用者環境に対応する。 ・利用者拡大のための機能を選択的に導入可能とする。 ・利用サービス向上のための機能を選択的に導入可能とする。

4.2.2. 「標準仕様書の考え方」の改訂案

本項は、現行の標準仕様書における要点を整理したものである「標準仕様書の考え方」に対して、調査検討結果を踏まえてどのような点を改訂するか、案を整理したものである。

(1) システム環境

(ア) システム構成

現行の標準仕様書は、オンプレミス型を前提としているが、サーバの高性能化や仮想化をはじめとするクラウド技術の進展により、情報システムにおけるクラウドの普及が進んでいる。クラウドを前提とした構成は、初期構築コストの軽減、インフラ運用負担の軽減が期待でき、予算と人員に制約がある公文書館等においては有益な選択肢となりうる。一方で機密性の高いデータを配置する場合、クラウドサービスのセキュリティ要件や機関のセキュリティポリシー等について十分検討が必要となる。

オンプレミス型、クラウド技術を活用した共同運用型はいずれも特徴や留意点があるため、各機関の事情に応じて柔軟に選択できるようシステム構成を例示する。

- 【必須】各機関の要望に応じた、システム構成や運用形態を選択可能とする。

(イ) サーバ、ネットワーク環境

サーバ、OS、ミドルウェアは、低価格化とオープンソース化が進んでいる。サーバ機器の性能向上に合わせて仮想化技術も発展しており、サーバ環境の整備におけるコスト面でのハードルは低減しているものと考えられる。

ネットワーク環境では、光回線・高速モバイル回線の普及が進んでおり、デジタルアーカイブ・システムで大容量のデジタルコンテンツを配信できる環境が整っている。

各機関が公開するデータ量、コンテンツ形式は様々であるため、各機関の事情に応じた構成を選択できるようにすべきである。

- 【必須】各機関の予算や規模(想定するアクセス数/トラフィック量)に応じた回線を導入するものとする。

(2) 目録情報のデータ化

(ア) 目録情報の項目

各機関で主に取り扱う資料区分には公文書(行政文書)、古書・古文書、刊行物、図書等がある[132]。目録情報の項目は、ISAD(G)等を参考に各機関の所蔵資料等の実情に合わせて整備する必要がある。その一方で、横断検索においては、複数機関間で共通した項目を用いる必要が生じる。

現在の標準仕様書では、Dublin Core の要素を参考に、横断検索のために必要な最小限の目録データの項目として、「ID」「年代(作成年度)」「資料名」「作成者(部署名)」「備考等」の 5 項目を設定している。項目が簡潔であることから多様な領域に適用できるが、情報の精密さが失われ、相互運用性を低下させ

---

[132] 国立公文書館 統括公文書専門官室 「平成 27 年度 歴史的資料として重要な公文書等の所在情報に関する調査報告書」(平成 28 年 3 月)

る可能性がある。

横断検索における利用者への提供情報の充実を図るためには、Dublin Core の基本 15 項目を拡張した DC タームスを用いる余地を残しておくことが望まれる。また、アプリケーションプロファイルの考えを取り入れることで、項目定義にかかるコストの軽減をはかりつつ、相互運用性を高めることができる。

- 【推奨】アプリケーションプロファイルの枠組みを採用し、DC タームスを用いて目録データの項目を拡張できるものとする。

（イ）記述方式、電子符号化

アーカイブズでの符号化記録史料記述の国際標準である EAD に対応することで、国内外を問わず目録情報の交換の容易さや機械可読性が高まり、目録情報の利活用を拡大させる効果も期待できる。

EAD は最新版の EAD3 がデータの相互運用性や多言語対応を行っており、より有効な手段となっている。ただし、正式に公開されたのが 2015 年であり、EAD3 は今後実用化されると見込まれる。

- 【推奨】目録情報の記述方式として EAD に準拠することを推奨する。（EAD データの仕様は、EAD2002 または EAD3 を推奨）

(3) 目録情報のデータベース化

（ア）目録データの管理・登録機能

デジタルアーカイブ・システムで扱うデータは非構造化データ等多様化が進んでおり、このような状況における情報検索の手段として、全文検索への対応がある。全文検索に対応する仕組み（全文検索エンジン）には、高機能・高性能なものがオープンソースでも提供されている。また、全文検索機能を有する RDBMS を用いることにより、データ管理の効率化を図る方法も考えられる。

各機関で目録データを作成する際に、個人情報や業務上の秘密情報を含む可能性があり、公開・非公開を設定できることは必須の機能である。更に共同運用システムでは、他機関の目録データへのアクセスを制限するために、各機関が管理する目録データへのアクセス権を管理する必要がある。

- 【必須】非構造化データ等を検索できる全文検索エンジンを採用する。
- 【推奨】目録データへのアクセス権を管理する機能を設ける。

（イ）検索機能

デジタルアーカイブ・システムのキーワード検索機能、詳細検索機能は、資料探索の基本的な手段として最低限必要な機能である。また、資料の階層構成を選択しながら目録を検索できる階層検索機能は、利用者の利便性を高め、一定の利活用促進の効果があると考えられる。

一方で、デジタルアーカイブ・システムの利用者は、必ずしも所蔵された資料に精通しているわけではない。多様な利用者の利用を促進するために、基本的な検索機能に加え、辞書検索機能、キーワードの候補を提示するサジェスト機能、検索結果を分類して表示するファセット機能等の検索支援手段を提供することが考えられる。これらの検索支援機能の実装においては、想定する利用者に合わせた検討を要する。

- 【推奨】利用者の検索を支援する機能を実装することを推奨する。

（ウ）業務機能

利用請求記録の管理機能に対するニーズは高いが、利用請求にかかる業務は各機関で異なるため、導入の容易さには幅があるものと考えられる。管理すべき情報や、データ化の方法について検討が必要である。更に共同運用においては、システム化に先立って、各機関の業務プロセスを標準化が必要となるが、業務の標準化を行うことにより、効率化によるコスト低減や、他業務への人員配置等が見込めるものである。

また、レファレンス機能のシステム化に対するニーズも高い。オンラインで利用者からの問い合わせに対して回答を行い、その知識をデータベース化することは、利用者の利便性向上と職員業務の負担軽減という効果が期待できる。さらに共同運用により、複数の機関がレファレンス情報を共有するならば、さらなる利便性の向上につながることが期待できる。

- 【推奨】利用請求記録の管理機能を推奨する。
- 【推奨】利用者からの問い合わせをインターネット上で受け付け、回答できる機能(レファレンス業務の管理機能)を推奨する。

（エ）バックアップ機能

デジタルアーカイブ・システムで安定したサービスを提供するため、データのバックアップ機能は必須の機能である。各機関が保有するデータ量と、将来的なデータ量の増加予測に応じて、適切なバックアップ方式やメディアを検討し導入することが推奨される。

また、大規模災害等に備えたデータの遠隔地バックアップがサービス継続のために望まれるが、各機関のデータ特性や重要性に応じて必要性を検討する必要がある。

- 【推奨】バックアップを遠隔地に保管することが望ましい。

(4)　デジタルコンテンツの取扱い

（ア）動画・音声データ

デジタルアーカイブで扱うコンテンツの多様化を踏まえて、動画・音声の配信について検討する必要がある。動画や音声の内部の符号化方式(コーデック)は多様であり、技術革新のサイクルが速い。利用者の利便性には普及度を勘案する必要があるものの、長期運用性の観点から、比較的長期にわたり活用されてきた実績のある方式をサポートすることが考えられる。

- 【必須】公開用の動画や音声データは、比較的長期にわたり活用されてきた実績のある方式をサポートする。

（イ）デジタルコンテンツの登録・管理機能

画像データの配信は、国際標準の配信方式を採用し、利用者の利便性とシステム間の相互運用性を高めることが求められる。

また、動画・音声データは、画像データと比べて容量が大きくなる傾向がある。従って、ストリーミング

配信による効率的な配信を実現することが望ましい。

- 【推奨】画像データの配信方式として、国際標準の配信方式を推奨する。
- 【推奨】動画・音声データの配信方式として、ストリーミングによる配信を推奨する。

(5) インターネットにおける情報提供

(ア) アクセスの永続性保証

資料目録やデジタルコンテンツの情報へのアクセスを永続的に保証することはデジタルアーカイブ・システムにおいて非常に重要であり、その実現手段としてパーマリンクを用いることが推奨される。

パーマリンクを実現するには目録やデジタルコンテンツの識別子、URI の設計を行う必要がある。また、その一貫性や永続性を保証するためには、各機関でシステム更新を行う際に URI が維持されることを明示的に仕様として入れる必要がある。

- 【推奨】目録情報、デジタルコンテンツ情報に対するパーマリンクを付与することを推奨する。
- 【必須】システム更新に際して、アクセスの永続性を考慮する。

(イ) 情報発信に必要な機能

デジタルアーカイブにおける情報発信は、Web 上で一般利用者へ提供することから HTTP は必須である。HTTP の最新版である HTTP/2 への対応は、閲覧速度向上により利用者のストレス軽減という効果が期待できるため、デジタルアーカイブ・システムでも対応を推奨する。

- 【推奨】HTTP/2(IETF RFC 7540)による情報発信を推奨する。

(ウ) 横断検索機能

デジタルアーカイブ・システムでは、HTTP ベースの検索プロトコルである SRU/W、OpenSearch のいずれかを実装することが、相互運用性の向上という点からも推奨される。

- 【必須】SRU/W、OpenSearch により情報発信できるものとする。

(エ) 一括提供機能

目録や資料画像等のデータを Web ブラウザ上で個別に閲覧するだけでなく、汎用的な形式で一括してダウンロードできることは、データの利活用拡大を促すことが期待できる。

一方で、データ連携の強化には、横断検索だけでなく機械可読形式によるデータ提供の機能が必要である。機関連携の枠組みについては今後の議論が待たれるが、デジタルアーカイブでもそれに対応する仕組みとして Web API 等を提供することが推奨される。

- 【推奨】目録データやデジタルコンテンツをユーザー操作により一括で提供できる機能を推奨する。
- 【推奨】Web API による情報提供機能を推奨する。

（オ） 利用規約及びライセンス、権利関係情報の明示

利用者が安心してアーカイブ・システムが提供するデータを利用できるようにするためには、政府標準利用規約やクリエイティブ・コモンズといった標準へ準拠し、明示することが望ましい。

- 【推奨】利用規約及びライセンス、権利関係を明示する。

（カ） 利用者拡大のための機能

利用者拡大には、SNS を用いて利用者が容易に情報発信を行えるよう支援することが有効であるが、各機関の方針に準じる必要がある。

利用者によるメタデータ付与等への能動的参加を実現するには、それらを実現する機能の整備に加えて、利用者による作業を管理する運用の検討が必要である。

- 【推奨】利用者が SNS を用いて情報発信を容易に行える仕組みを推奨する。

(6) 非機能要件

（ア） セキュリティ対策

サイバー攻撃の手段は多様化・複雑化しており、デジタルアーカイブ・システムにおいても安定的なサービスの提供、データ保全のために十分なセキュリティ対策を講じる必要がある。システム上の対策として様々な選択肢があるが、セキュリティレベルに応じてコストが高くなる。各機関のセキュリティ対策実施状況や予算に応じて選択する必要があるが、最低限の対策としてウイルス対策ソフトと WAF(Web Application Firewall)の導入を推奨すべきである。また、デジタルアーカイブ・システムの構築事業者に対しては、IPA の示す基準に基づいたチェックを義務付けるべきである。

暗号化については、業務用データの取り扱いや利用者のセキュリティへ配慮することを踏まえて、HTTPS もサポートすることを推奨する。

- 【推奨】ネットワーク上のセキュリティに配慮したシステムとするため、ウイルス対策ソフト及び WAF(Web Application Firewall)の導入を推奨する。
- 【必須】システム構築・導入事業者に対して、情報処理推進機構(IPA)の示す「安全なウェブサイトの作り方」に準拠することを義務付ける。
- 【推奨】HTTPS（IETF RFC2818）による情報発信を推奨とする。

（イ） 性能要件

デジタルアーカイブにおける性能要件としては、利用者側の観点（目録の検索速度、デジタルコンテンツの配信・閲覧速度等)と業務側の観点（目録データの登録時間、デジタルコンテンツの登録時間等)の 2 つが重要となる。いずれも各機関のデータ量、人的資源、業務内容により求められるレベルは異なると考えられるため、求める性能要件を明確化することが必要である。

- 【必須】目録検索速度、デジタルコンテンツの配信・閲覧速度等の性能要件は、利用者が快適

に利用できる水準とする。

- 【必須】目録データの登録時間、デジタルコンテンツの登録時間等の性能要件は、業務に支障をきたさない水準とする。

（ウ）運用保守

運用保守については各機関の予算及び人的資源の確保が難しい状況であるため、各機関の事情を踏まえたシステム構成やアプリケーション機能を選択できることが望ましい。

- 【必須】各公文書館等の予算や規模に応じた運用保守が行えるシステム構成とする。