

研究論文

『文部省往復』を中心としたデジタルアーカイブの構築とその活用

中村覚 稗方和夫 満行泰河 加藤諭 宮本隆史 高嶋朋子

1. 諸言

東京大学文書館は、東京大学百年史編集室および東京大学史史料室で収集した資料及び成果を引き継ぐ組織として設置された。東京大学にとって重要な法人文書及び東京大学の歴史に関する資料等の適正な管理、保存及び利用等を行うことにより、東京大学の教育研究に寄与することを使命とした組織である。

東京大学文書館が所蔵する資料は大きく『特定歴史公文書等』と『歴史資料等』の二種類に分けられる。『特定歴史公文書等』とは、東京大学の教職員が職務上作成・収受し、組織的に用いてきた文書等（法人文書等）のうち、歴史資料として重要なものとして東京大学文書館に移管された資料である。『歴史資料等』は、歴史的・文化的な資料、または学術研究用の資料として、東京大学に関係する団体や個人等から収集した資料である。

前者の『特定歴史公文書等』に含まれる資料の一つに『文部省往復』がある。文部省往復とは、東京大学と文部省との間でやりとりされた公文書綴である。具体的には、東京大学の前身である南校において記録・保存された日誌や諸省とのやり取りの記録、職員録などが含まれる。2013年2月に重要文化財指定を受け、学術的に重要な資料として評価を得ており、研究資料として実際に用いられている [1] [2]。

東京大学文書館では、科学研究費プロジェクト¹を通じ、明治期分137簿冊分の文部省往復のデジタル画像化・メタデータ作成作業を完了させている。また、この成果を東京大学文書館のHP[3]にて公開している。ただし、現在は簿冊毎のメタデータを表形式で表示するWebページを設け、対象簿冊の画像をまとめたPDFファイルとともに提供する公開方法を採用している。このため、「簿冊を横断した資料検索ができない」「画像とメタデータ（目録情報）を人手で紐づける必要がある」などの課題が存在する。

本研究では、上記の課題解決に向け、文部省往復を対象としたデジタルアーカイブのプロトタイプ²の構築を一つの目的とする。「デジタルアーカイブ」という用語には様々な解釈が存在するが、本研究ではデジタル化された歴史資料に関する目録情報をデータベース化し、目録情報に基づく資料検索や閲覧が可能なシステムとする。また、このデジタルアーカイブのプロトタイプ構築において、近年データの共有や再利用の容易性の観点から注目されているLinked Dataを用いることにより、単なる資料公開に留まらず、その活用支援を目的とする。本研究では、この具体的な活用例として、資料研究やデジタル展示への事例適用を通じ、Linked Dataを用いたデジタルアーカイブの有用性について検討する。

なお、本稿は筆者の個人的見解、およびそれに基づく取り組み内容を述べたものであり、組織としての公式見解を示すものではない点に注意されたい。

2. 関連研究

デジタルアーカイブは先述したように厳密な定義を持たないが、一般的には図書・出版物、公文書、美術品・博物品・歴史資料等をデジタル化し、インターネット上で共有・利用できる仕組みとして理解される。デジタルアーカイブ化により、資料の半永久的な保存、および時間的・地理的な制約を超えた資料提供が可能となる。一方、今日のデジタルアーカイブに共通する課題として、これまでは資料の保存と公開に主眼が置かれ、その利用方法に関する検討が十分になされてこなかった点が挙げられる。この課題に対し、馬場ら [4] はデジタルアーカイブの意義をその作成自体に置くのではなく、歴史学研究における利用の観点から考えるべきであると述べ、研究成果の公開手段・技術としての活用を提案している。また国立歴史民俗博物館では、「資源」「研究」「展示」という三つの要素を有機的に連鎖させ、さらに積極的に「共有・公開」することによって、博物館機能を存分に発揮する独自の研究スタイル「博物館型研究統合 [5]」を提唱、実践している。

このデジタルアーカイブの活用に向け、近年では欧米を中心として Linked Data の活用が進められている [6]。Linked Data とは、Web 上で機械可読な構造化データを公開する技術の総称であり、分野やシステムを超えたデータ共有や再利用が容易となる利点を持つ [7]。Linked Data 形式でデジタルアーカイブのデータを公開することにより、複数のデジタルアーカイブ間の連携や、第三者によるアプリケーション開発が容易となり、再利用・再配布による新たな価値の創造が期待されている。

Linked Data の主要な技術的要素として、URI (Uniform Resource Identifier) と RDF (Resource Description Framework) が挙げられる。URI は、URL (Uniform Resource Locator) の考え方を拡張した識別子であり、Web 空間の「もの (リソース)」を一意指し示すものである。この URI はそれを作成した機関が提供するデータセットの中だけでなく、他の機関が作成するデータセットの中でも使用することができる。この仕組みにより、データセットを超えて相互に参照可能な Linked Data が可能になる。また、これらのデータは主に RDF という標準言語を用いて記述する。RDF では、主語、述語、目的語の3つの要素で「もの」に関する関係情報を表現し、これを RDF トリプルと呼ぶ。この RDF トリプル中における「もの」を URI によって表現することにより、ある RDF トリプルの目的語で使われた URI と別の RDF トリプルの主語で使われた URI が同一であれば、これらの RDF トリプルは URI を介して関連づけられる。この RDF によるデータ記述の例を図 1 に示す。

図左部は、URI 「<http://utarchives/XXXX/material>」によって表現される東京大学文書館の一所蔵資料について、表題が「大学ニ四学部ヲ置ノ件」、差出人が「田中不二麿」であるという目録情報を RDF によって記述している例を示す。この時、差出人「田中不二麿」を単なる文字列ではなく、「田中不二麿」を示す URI 「<http://ja.dbpedia.org/resource/田中不二麿>」を用いて表現することにより、対象 URI を用いて情報を記述する他のデータベースとの関連づけが可能となる。図右部は、あるデータベースが「田中不二麿」に関するサムネイル画像や説明文を RDF によって記述している例を示す。本図は「田中不二麿」を示す URI を介して、

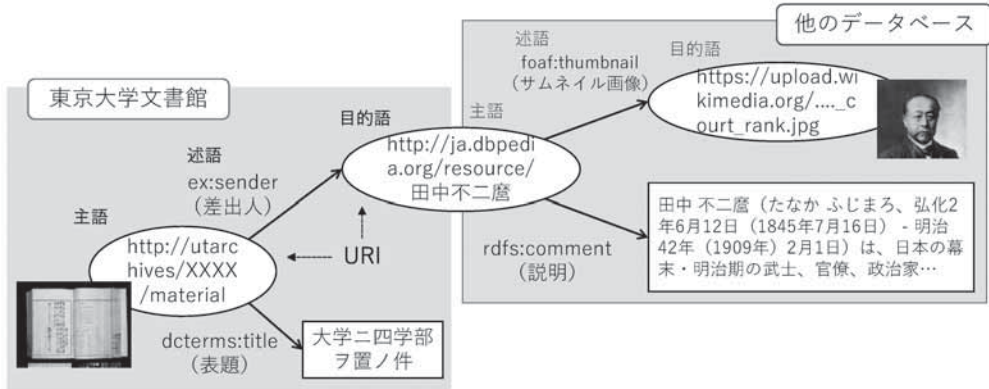


図1 RDFによるデータ記述例

異なるデータベースにおける情報が関連づけられている例を示している。

この Linked Data の活用例として、欧州 35ヶ国、3,000 以上の図書館・美術館・博物館・文書館等の文化施設が保有する資源へのアクセスを可能とする横断検索ポータルサイト「Europeana」が運用されている [8]。また、このような複数の文化施設を横断した検索の実現に向け、目録情報を記述するためのスキーマの標準化が進められている。例えば、アメリカを中心として世界各国の大学や研究機関で構成された機関である OCLC (Online Computer Library Center) は、典拠データや各種コード類の Linked Data をサービスとして実現している [9]。さらに、国立国会図書館では、情報資源の組織化および利用提供のためのメタデータ記述の標準として「国立国会図書館ダブリンコアメタデータ記述 (DC-NDL)」を定めている [10]。久山ら [11] は浮世絵が所蔵されている複数の異種データベースに対して、Linked Data を用いて多言語統合アクセスを実現し、複数の異種浮世絵データベース間における同一作品の同定手法を提案している。また、林ら [12] は Linked Data を活用し、人文系データベースの相互運用を考慮したリポジトリシステムを開発し、その適用事例として先祖由緒并一類附帳データベースを構築している。さらに Kiourt [13] らは Web 空間上でバーチャル展示を可能とするシステム「DynaMus」を開発している。DynaMus で用いる展示資料の情報源の一つとして、Europeana が公開する Linked Data を用いることにより、資料形態をはじめとする目録情報を利用している。

本研究では、東京大学文書館の所蔵資料の一つ『文部省往復』を対象としたデジタルアーカイブのプロタイプ構築において、上述した Linked Data を活用する。これにより、単なる資料公開を目的としたシステム構築に留まらず、公開資料の活用を支援する。また、資料研究やデジタル展示への適用を通じ、デジタルアーカイブおよび Linked Data の具体的な活用方法を示すことが本研究の目的である。

3. 提案手法

3.1. 提案手法の全体像

本研究で提案するデジタルアーカイブの構築および活用手法の全体像を図2に示す。まず、RDF トリプルを格納するためのデータベースである「RDF ストア」を Web 空間上に構築する。この RDF ストアを用いて、資料の目録情報をはじめとする「目録データ」、研究者が研究目的に応じて蓄積する「研究データ」、キャプションやキーワード等の資料展示に必要となる「展示データ」を一元的に管理する。さらに、デジタル化した一次資料を格納するデータベースや、後述する DBpedia などの他の RDF ストアと URI を介して連携する。この RDF ストアに対して、RDF データに対するクエリ言語である SPARQL 等の API を利用することにより、アプリケーション構築等によるデータ活用を行う。本研究では、このアプリケーションの一つとしてデジタルアーカイブを位置づけるとともに、資料研究や展示による成果公開などの活動におけるデータ活用を支援する。

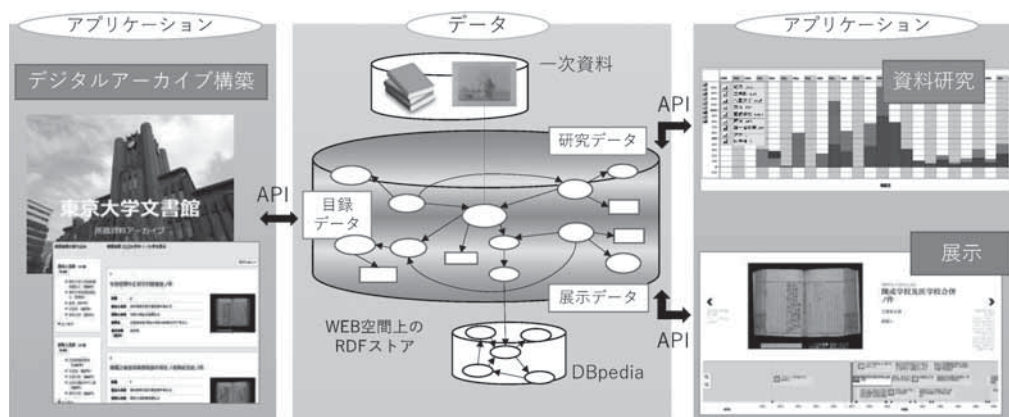


図2 提案するデジタルアーカイブ構築および活用手法の全体像

3.2. RDF によるデータ記述

ここでは RDF ストアに格納される RDF データの記述について述べる。この記述例を図3に示す。各資料に一意となる URI を割り当て、これを主語とする述語と目的語をメタデータとして記述する。これにより、資料の目録情報をはじめとする「目録データ」や、研究者が研究目的に応じて蓄積する「研究データ」、キャプションやキーワード等の資料展示に必要となる「展示データ」を一元的に管理する。以下では、これら三種類のデータの各々について説明を加える。

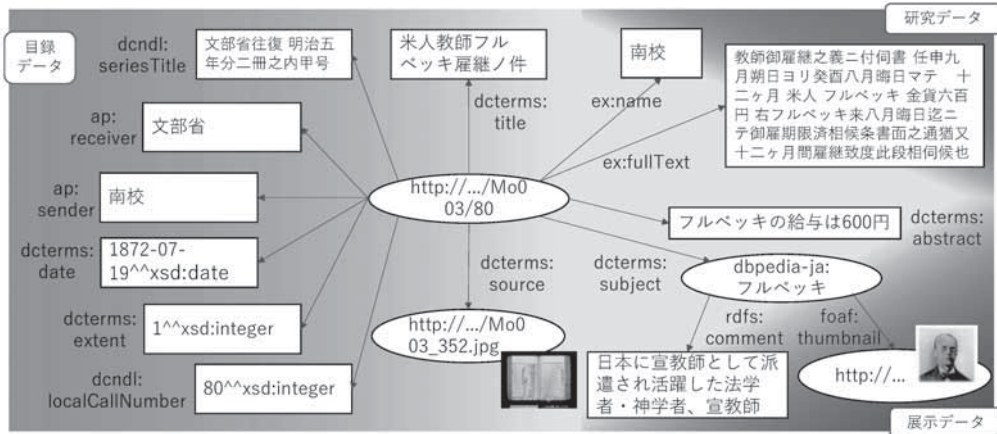


図3 構築したRDFデータの例

3.2.1. 目録データ

目録データは、図3左部に示すような目録情報である。『文部省往復』では、資料名や作成日などの一般的な目録情報に加え、往復文書の特徴である「差出人名前」や「受取人名前」などの情報が記載されている。RDFでは述語をURIによって記述するため、これらの資料名や作成日、「差出人名前」や「受取人名前」などのメタデータ項目毎にURIを割り当てる。このメタデータ項目とURIのマッピング結果について、主要なものを表1に示す。

表1 主要なメタデータ項目とURIのマッピング

	URI	ラベル	値の例
1	dcterms:title	タイトル	米人教師フルベッキ雇継ノ件
2	dcterms:extent	数量	1^^xsd:integer
3	dcterms:date	作成日	1872-07-19^^xsd:date
4	dcterms:source	画像資料へのパス	http://.../Mo003_352.jpg
5	dcndl:localCallNumber	整理番号	80^^xsd:integer
6	dcndl:seriesTitle	簿冊名	文部省往復 明治五年分二冊之内甲号
7	ap:sender	差出人名前	南校
8	ap:receiver	受取人名前	文部省

資料名や作成日等の一般的なメタデータ項目については、目録情報を記述するための標準的な語彙を提供する Dublin Core[14] を用いる。資料の整理番号、および簿冊名を示すメタデータ項目に対しては、国立国会図書館が提供する DC-NDL[15] を使い、それぞれ「dcndl:localCallNumber」、「dcndl:seriesTitle」をマッピングする。また、「差出人名前」や「受取人名前」などの既存の語彙では提供されていないメタデータ項目については、独自に定義した名前空間「http://apps.is.k.u-tokyo.ac.jp/utarchives/property/」を定義し、接頭辞として「ap」を用いてそれぞれ「ap:sender」、「ap:receiver」を与えた。さらに、各資料に関する画像資料については、画像の URL に基づく URI を与え、対象画像資料を目的語とするメタデータを「dcterms:source」を用いて関連づける。

3.2.2. 研究データ

図3右上部に示す研究データは、研究者が各々の研究目的に応じて整理するデータである。これらの情報も上記の目録データと同様に、資料のメタデータとして管理する。この時、研究に必要な情報は研究目的に応じて様々であるため、研究者が任意にメタデータ項目を定義し、必要な情報をその値として蓄積する。例えば図3では、翻刻結果を管理するためのメタデータ項目「ex:fullText」を定め、その文字列情報をメタデータとして管理している例を示す。

3.2.3. 展示データ

最後に図3右下部に示す展示データについて述べる。展示においては、歴史になじみのない一般の利用者が対象となる。このため、展示対象となる資料には一般的な目録情報に加え、興味や関心を喚起させる資料説明（キャプション）やキーワードの整理、地図や年表などを用いた視覚的な情報提供が求められる。これらの情報に対し、本研究では Dublin Core で提供されている語彙を用い、資料のメタデータとして記述する。キャプションについては情報資源の要約や抄録等の記述に用いる「dcterms:abstract」、キーワードについては情報資源の内容を表す統制語彙やフリーキーワード等を URI によって表現するために用いる「dcterms:subject」を用いて記述する。特にキーワードについては、Wikipedia から情報を抽出して Linked Data として公開するプロジェクトである「DBpedia [16]」を活用する。DBpedia が提供する URI を値として与えることにより、本研究で作成するデータと DBpedia が提供するデータが URI を介して関連づけられる。これにより、対象キーワードの概要説明やサムネイル画像を利用することが可能となり、一般利用者への専門用語や背景知識に関する情報の補足を支援する。

4. 構築したデジタルアーカイブのプロトタイプ

ここでは、3.2.1 で述べた目録データを用い、構築したデジタルアーカイブのプロトタイプについて述べる。先述したように、本システムは Web 空間上の RDF ストアで管理されているデータに対し、API によってアクセスする Web アプリケーションである。RDF ストアとしては「Virtuoso [17]」を使用し、アプリケーション構築には Java を用いた。

本システムは、先述した科学研究費プロジェクト¹を通じて、デジタル画像化・メタデータ作成作業を完了している明治期分 137 簿冊分の文部省往復、具体的には明治 4 (1871) 年から明治 45・大正元年 (1912) 年にかけて作成された 11,367 件の資料を対象とした。

構築したデジタルアーカイブのプロトタイプを検索インタフェースを図 4 に示す。本インタフェースの画面右部には、各資料の表題やサムネイル画像などの一覧が表示される。また画面左部には、検索結果として得られたメタデータ毎の資料数の集計結果が表示される。

The screenshot displays a search interface for a digital archive. On the left side, there are three filter panels:

- 画像 (Image):** A search box containing 'あり (11,367 件)' and a search button.
- 差出人名前 (Sender Name):** A list of 1,410 items with 30 displayed. The list includes:
 - 東京大学三学部総理加藤弘之 (904 件)
 - 東京大学総理加藤弘之 (515 件)
 - 総長 (511 件)
 - 文部省 (367 件)
 - 東京大学 (331 件)
- 受取人名前 (Receiver Name):** A list of 1,281 items with 30 displayed. The list includes:
 - 文部省編同僚第 (1,057 件)
 - 文部省 (847 件)
 - 文部大臣 (406 件)
 - 文部大輔田中不二雄 (369 件)
 - 東京大学 (262 件)

On the right side, the search results are displayed in a list format:

- Item 1:**
 - 表題:** 年度経費中正貨交付額増加ノ件
 - 階層:** 所蔵資料 > 特定歴史公文書等 > 01 事務 > 108 総合企画 > S0001 文部省往復 > Mo046 往復簿 明治十四年 丁二
 - 数量:** 2
 - 差出人名前:** 會計局長文部大書記官中島永元
 - 受取人名前:** 文部三等出仕加藤弘之
 - 簿冊名:** 文部省往復 明治十四年分四冊之内丁号之二
 - 目次分類 (部;件):** 会計局
- Item 2:**
 - 表題:** 廃職之後従前事務取扱中再任ノ者俸給支給ノ件
 - 階層:** 所蔵資料 > 特定歴史公文書等 > 01 事務 > 108 総合企画 > S0001 文部省往復 > Mo046 往復簿 明治十四年 丁二
 - 数量:** 1
 - 差出人名前:** 會計局長文部大書記官中島永元
 - 受取人名前:** 東京大学総理加藤弘之
 - 簿冊名:** 文部省往復 明治十四年分四冊之内丁号之二
 - 目次分類 (部;件):** 会計局

Each result item includes a thumbnail image of the document.

図 4 構築したデジタルアーカイブの検索インタフェース例

その他、表題や差出人名前、作成日等のメタデータに基づく検索インタフェースや、各資料のメタデータの一覧を閲覧するための詳細インタフェース、画像資料の拡大・縮小・回転が可能な閲覧インタフェースを提供する。

これらのインタフェースを用いることにより、簿冊を横断した資料の検索や、画像と目録データに基づく資料の閲覧が可能となる。この結果、1 で挙げた既存の公開方法の課題を解決するデジタルアーカイブのプロトタイプを構築することができた。

5. 活用事例

ここでは、4で構築したデジタルアーカイブのプロトタイプの活用事例について述べる。具体的には、Linked Dataとして提供される目録データを用い、『文部省往復』を用いた資料研究事例と、資料のキャプションや年表等を用いてWeb空間上で資料展示を行うデジタル展示事例について述べる。

5.1. 資料研究事例への活用

5.1.1. 研究課題

1で述べたように、『文部省往復』は明治初期の「大学」の設立過程を検討する上で重要な資料である。本研究事例ではこのような資料の特徴、および11,367件の資料に関する大規模なメタデータを利用可能な利点を生かし、明治初期における東京大学の設立史の全体像を俯瞰することを目的とする。具体的には、デジタル画像化・メタデータ作成作業が完了している文部省往復を用い、明治4(1871)年から明治19(1886)年の帝国大学令の公布に至るまでの組織名称の変遷史の可視化を行う。

5.1.2. 研究データの整理

東京大学の設立史の俯瞰に向け、本研究では資料のメタデータとして整理された「差出人名前」と「受取人名前」を利用する。これらのメタデータに出現する東京大学の組織名称に基づいて資料を時系列に並び替え、東京大学の名称の変遷史を可視化する。これらのメタデータについて、例えば東京大学が法文理の三学部によって構成されていた時期では、「東京大学三学部」や「東京大学法理文三学部」などの表記揺れが存在し、資料の比較や分類に適さない形式で記述されている。したがって、これらのメタデータから比較対象とする組織名称を定め、これらの名称に基づく資料分類を目的とした研究データとして新たに付与した。具体的には組織名称「南校」「東校」「第一番中学」「開成学校」「医学校」「三学部」「帝国大学」、および組織の名称によって変化する役職名称「総理」「総長」の計九つを対象とし、これらの名称をメタデータ項目「ex:name」を用いて新たに付与した。なお、「東京大学」という名称については、時期に関わらず恒常的に出現するため、本研究においては対象外とした。

5.1.3. 資料分析

5.1.2で整理したメタデータに基づいて資料を分類し、その結果を時系列に並び替えて可視化した結果を図5に示す。この可視化図は、RDFストア内の目録データの一つである資料の「作成日」、および研究データとして新たに追加した「組織名称」をSPARQL APIによって抽出し、この抽出結果を可視化した結果である。

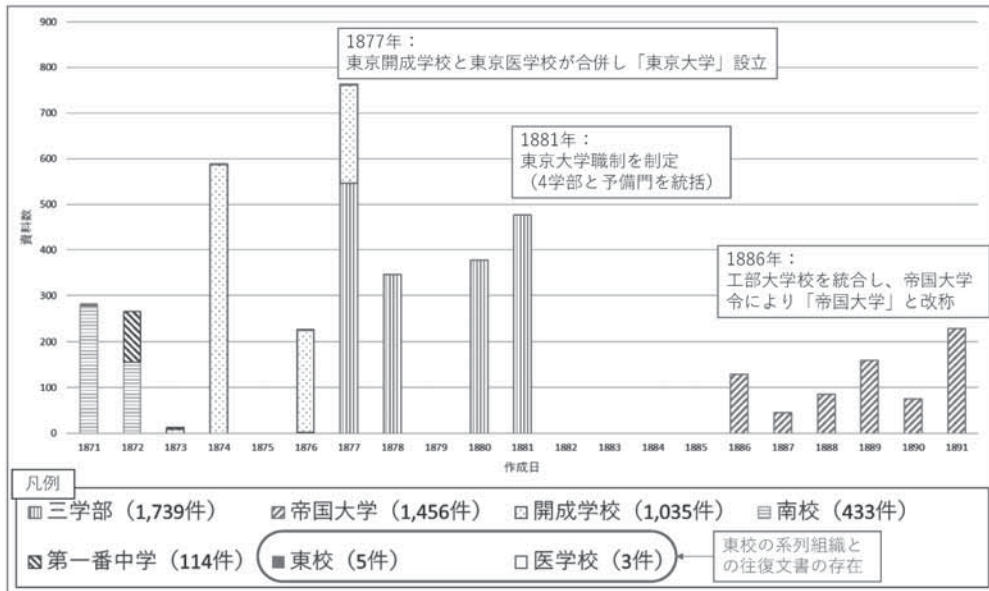


図5 明治初期における東京大学の名称の変遷史の可視化

この結果、明治10(1877)年4月の東京大学創設以前には、「南校」や「第一番中学」「開成学校」という名称が使用されていたのに対し、それ以後は「三学部」という名称が使用されている。また、明治14(1881)年6月の東京大学職制の制定以後は「東京大学」という名称のみが使用されるため、本可視化結果においては資料数が0件となっている。さらに、明治19年(1886)年3月の帝国大学令の公布後は「帝国大学」という名称が用いられていることがわかる。同様に、「役職名」に関する資料の出現頻度の可視化を行った結果、1877年以降は役職名「総理」が使用されているのに対し、1886年以降は「総長」に変化していることを確認した。これらの結果から、本可視化図が東京大学の設立史を正しく表現していることが確認できた。以上のように、本研究事例では『文部省往復』のメタデータを用いることにより、東京大学の組織名称の変遷に関する全体像を俯瞰することができた。

さらに、玉井 [1] は『文部省往復』の特徴を以下のように述べている。

“(文部省往復)に東校、医学校、東京大学医学部と続く組織に関する文書は収められていない。これは南校および東校の系列が、明治10(1877)年に東京大学として合併して以降も、それぞれ別組織として運営がなされていることに依拠しており、明治14(1881)年に別組織を統合する総理が創設されるまで続いた。”

この記述に対し、図5下部の凡例部に示したように、「東校」との往復文書5件、「医学校」との往復文書3件の存在を確認した。「東校」との往復文書の例としては、明治4(1871)年に作成された簿冊「文部省及諸向往復 明治四年之分式冊ノ内乙号」の目次分類「諸学校往復

之部」に含まれる文書「亜硫酸御廻被下度旨ニテ東校へ往書」や「動物図式ノ件ニ付東校エ往復書」が挙げられる。また、「医学校」との往復文書については、「文部省」から「開成学校」と「医学校」の両者に合わせて送付された文書「諸官省達回付ノ件」（明治7年）や「経費金額改定ノ件」（明治10年）であった。

東校系列の組織と文部省の往復文書については、『東京帝国大学五十年史料』[18]に別系統として存在するものではあるが、本資料研究事例を通じ、『文部省往復』にも少数ながら存在するという新たな特徴を見出すことができた。このような新たな特徴を発見することができた要因として、デジタルアーカイブが提供する大規模な目録データの活用による定量的な分析が可能となった点が挙げられる。この点について、本資料研究事例の共同研究者から、「研究対象資料および研究領域に高い専門性を有する研究者が無意識的に資料群に対する理解を作り上げてしまうことがある点に対して、定量的な分析や可視化が新たな気づきを提供することができる点で有用である」という評価を得た。これは資料研究プロセスにおいて構築したデジタルアーカイブを効果的に活用できたことを示し、本研究で提案した手法の有用性を支持する結果である。

5.2. デジタル展示事例への活用

構築したデジタルアーカイブの活用例の二点目として、資料のキャプションや年表等を用いて Web 空間上で資料展示を行うデジタル展示事例について述べる。

5.2.1. 展示テーマ

本展示事例では、東京大学広報室が2014年7月25日に発行した『学内広報』に掲載された東京大学文書館の所蔵資料16点を対象とする。各資料には原資料の記載内容に基づく目録情報だけでなく、一般の利用者の興味や関心を喚起させるためのキャプションとともに資料が掲載されている。例えば、目録データとして表題「米人教師フルベッキ雇継ノ件」が与えられている資料に対し、展示資料名として「フルベッキの給与は600円」、キャプションとして「お雇い外国人の給与は一人一人異なっており、その都度、文部省に伺いをたてて認めてもらう必要がありました。・・・」などの情報が付加されている。本事例では、3.2.3で述べた記述手法に基づき、これらの展示データをRDFによって記述し、目録データと組み合わせたデジタル展示を行う。

5.2.2. 構築したデジタル展示システム

デジタル展示に向け、資料のキャプションや年表等を用い、Web空間上で資料展示を行うシステムを構築した。本システムは4で構築したデジタルアーカイブと同様、RDFストアに対してAPIによってアクセスし、データを利用するWebアプリケーションである。目録データとして管理されている資料の画像情報と作成日、および新たに追加した展示データをSPARQLによって取得し、システム上で可視化している。この開発したデジタル展示システムのインタフェース例を図6に示す。

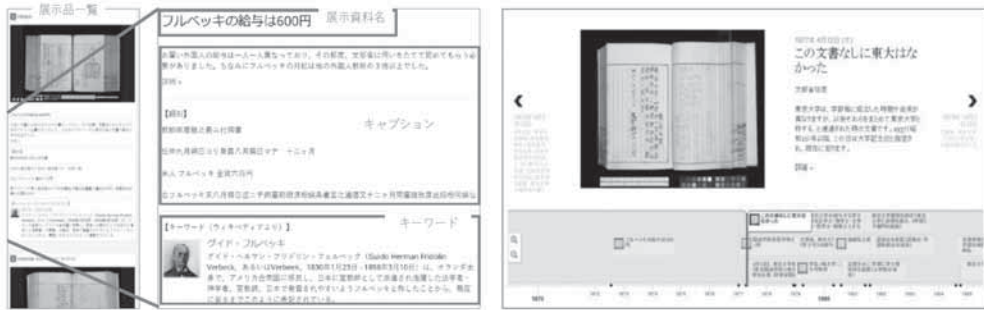


図6 展示資料の一覧表示(左)とタイムライン表示(右) インタフェース例

図左部に示すインタフェースは展示資料を一覧形式で表示し、各展示資料には資料名やキャプション、および資料に関連するキーワードを表示する。このキーワードに関しては3.2.3で述べたように、DBpedia からサムネイル画像や概要説明を自動的に取得し、表示している。

また図右部は資料の作成日を利用し、年表形式で表示するインタフェースである。展示資料だけでなく、東京大学の歴史に関する年表 [19] を合わせて表示することにより、資料の時代背景等を加味した情報提供を行う。

本システムは東京大学柏キャンパス一般公開(2016年10月21日から22日)に合わせて東京大学文書館が企画した展示会において利用され、原資料の展示と並行し、本システムを用いたデジタル展示を実施した。展示企画者からは、本郷キャンパスなどの他の館に所蔵されている資料の展示も可能となった点など、本システムに対する良好な評価を得た。

6. 考察

本稿では、東京大学文書館が所蔵する『文部省往復』を中心としたデジタルアーカイブのプロトタイプ構築、およびその活用例としての資料研究事例、デジタル展示事例について述べた。以下では、これらの事例毎に提案手法の有用性の考察と今後の課題について述べる。

6.1. デジタルアーカイブ構築

デジタルアーカイブ構築に関しては、Linked Data を用いたデータの共有や再利用の支援を目的とし、公開データを API を用いて利用するアプリケーションとして構築した。これにより、資料公開を目的としたデジタルアーカイブ構築だけでなく、資料研究やデジタル展示への活用が可能となった。さらに、これらのプロセスで新たに整理した研究データや展示データを目録データとして活用することにより、デジタルアーカイブの改良につなげることができた。例えば、キーワード(例「ガイド・フルベッキ」など)に関する情報を含む展示データについて、デジタルアーカイブが発行する SPARQL クエリの検索条件に当該データを加えることにより、

目録データとしては未整理であったキーワードに基づく資料検索が可能となった。このように、デジタルアーカイブが提供する目録データをデジタル展示に活用し、さらにこの活動成果をデジタルアーカイブの改良に応用できた点は、Linked Data の導入によるデータの利活用が実現した例であり、本研究の有用性を示す一つの結果である。

一方、東京大学文書館では『文部省往復』以外にも、『東京帝国大会計課文書』や元総長・元学生の資料などの多様な資料を所蔵しており、資料を作成した組織やその階層構造に基づく資料管理などが求められる。このような『文部省往復』以外の資料を含めた資料管理に向け、今後はオーストラリア国立公文書館などが採用する「シリーズ・システム [20]」などを導入し、資料管理体系の RDF 表現についても検討していく。

6.2. 資料研究事例

資料研究事例においては、『文部省往復』の目録データを利用し、明治初期の東京大学の設立史に関する資料研究を行った。RDF ストアに対する検索 API の活用により、資料分析における目録データの活用が容易となり、特に大規模なデータを用いた定量的な分析が可能となった。この定量的な分析に基づき、明治初期における東京大学の名称の変遷史の可視化や、分析対象資料に関する新たな特徴を見出すことができた点は、資料研究に対する有用性を示す結果である。

一方、本事例では資料のメタデータに基づく定量的な分析に留まった。今後は資料の内容に基づく定性的な分析や、組織の関係性に基づく資料研究への応用を行う。

6.3. デジタル展示事例

展示事例では、Web 空間上で資料展示を行うデジタル展示を行った。Linked Data を用いることにより、目録データと展示データを関連づけて管理することが可能となり、デジタル展示を目的としたアプリケーション開発へのデータ活用が容易となった。また、DBpedia のような外部の Linked Data と関連づけることにより、専門知識や背景知識などのキーワードに関する情報を半自動的に拡充した。これにより、一般的な目録情報だけではなく、歴史になじみのない一般の利用者を対象としたデータ作成が容易となった。

なお、本適用事例では過去に整理した展示データの再利用によるデジタル展示に留まった。今後は、展示の企画プロセスへの活用を検討し、展示企画におけるストーリーラインの作成や、展示資料の選定などへの応用を行う。

7. 結 論

本稿では、東京大学文書館が所蔵する『文部省往復』を中心としたデジタルアーカイブのプロトタイプ構築、およびその活用例としての資料研究事例、デジタル展示事例について述べた。デジタルアーカイブのプロトタイプ構築については、11,367 件の資料に対する検索や閲覧が可能なシステムを構築した。また、近年の潮流である Linked Data を用いてデジタルアーカイ

ブを構築することにより、資料公開だけでなく、資料活用を支援する環境を実現した。この結果、活用例の一つである資料研究事例においては、『文部省往復』の目録データ中に登場する組織名称に基づいて資料の出現頻度を可視化し、明治初期における東京大学の名称の変遷史を俯瞰することができた。また、「東校」系列の組織との往復文書の存在を確認し、『文部省往復』における新たな特徴を見出すことができた。さらに、活用例の二点目であるデジタル展示事例においては、『文部省往復』を含む東京大学文書館の所蔵資料16点をWeb空間上で公開するデジタル展示システムを構築し、展示スペースの問題や資料の所在地に関する制約を超えた資料展示を実現した。

なお、本稿はデジタルアーカイブとLinked Dataの技術的活用における実現可能性の検証を目的とした。今後は、デジタルアーカイブの運用面における課題整理などを通じ、持続可能性のあるシステムの在り方について検討していく。

註

- 1 「文部省往復を基幹とした近代日本大学史データベース」(吉見俊哉教授代表・課題番号15HP8023)
- 2 試作品。本稿では、動作や機能を検証するために試作されたソフトウェアを意味する。

参考文献

- [1] 玉井建也「『文部省往復』から見る明治初期の『大学』成立過程」『東京大学史紀要』第21号、2014年、1-13頁。
- [2] 頼住一昭「『文部省往復』にみる『剣術柔術調査』の諮問委員の人選に関する一考察(体育史)」『日本体育学会大会号』第49号、1998年、147頁。
- [3] 特定歴史公文書等 文部省往復 | 東京大学, Available at: <<http://www.u-tokyo.ac.jp/history/S0001.html>>. Accessed on: Nov 10th 2016.
- [4] 馬場章、吉田成、谷昭佳、吉田正高、吉仲亮、川瀬敏雄、肥田康、吉谷隆彦、津田光弘「デジタルアーカイブからデジタルエキジビジョンへ」『じんもんこん2001論文集』第2001巻、2001年、17-24頁。
- [5] 国立歴史民俗博物館、重信幸彦、小池淳一編『民俗表象の現在：博物館型研究統合の視座から』岩田書院、2015年。
- [6] 橋詰秋子「なぜ図書館はLinked Dataに取り組むのか欧米の事例から」『情報管理』第58巻第2号、2015年、127-134頁。

- [7] ヒース、トム他著、武田英明他訳『Linked Data : Web をグローバルなデータ空間にする仕組み』近代科学社、2013年、7頁。
- [8] Europeana Creative Challenges, Available at: <<http://pro.europeana.eu/web/europeana-creative/challenges>>. Accessed on: Nov 10th 2016.
- [9] T. Fons, P. Jeff, W. Richard, OCLC's Linked Data Initiative, 'Using Schema.org to Make Library Data Relevant on the Web', *Information Standards Quarterly*, vol. 24, no. 2/3, 2012, pp. 29-33.
- [10] 中山正樹「国立国会図書館におけるデジタルアーカイブ構築 知の共有を目指して」『情報管理』第54巻第11号、2011年、715-724頁。
- [11] 久山岳夫、Batjargal Biligsaikhan、木村文則、前田亮「複数の異種浮世絵データベース間における同一作品の同定手法の提案」『じんもんこん 2013 論文集』第2013巻第4号、2013年、225-232頁。
- [12] 林正治、堀井洋、堀井美里、宮下和幸、中野節子、山地一禎、高田良宏「リポジトリシステムを利用した先祖由緒并一類附帳データベースの構築」『じんもんこん 2012 論文集』第2012巻第7号、2012年、239-246頁。
- [13] C. Kiourt, A. Koutsoudis, G. Pavlidis, 'DynaMus: A Fully-Dynamic 3D Virtual Museum Framework', *Journal of Cultural Heritage*, vol.22, 2016, pp. 984-991.
- [14] ダブリンコア (Dublin Core), Available at: <<http://www.ndl.go.jp/jp/aboutus/standards/translation/dcmi-terms.htm>>. Accessed on: Nov 10th 2016.
- [15] 国立国会図書館ダブリンコアメタデータ記述 (DC-NDL), Available at: <<http://www.ndl.go.jp/jp/aboutus/standards/translation/dcmi-terms.htm>>. Accessed on: Oct 10th 2016.
- [16] DBpedia Japanese, Available at: <<http://ja.dbpedia.org/>>. Accessed on: Nov 10th 2016.
- [17] OpenLink Virtuoso Home Page, Available at: <<http://virtuoso.openlinksw.com/>>. Accessed on: Nov 10th 2016.
- [18] 東京帝国大学著『東京帝国大学五十年史』東京帝国大学、1932年。
- [19] 沿革 | 東京大学, Available at: <http://www.u-tokyo.ac.jp/gen03/b03_01_j.html>. Accessed on: Nov 10th 2016.
- [20] 国文学研究資料館編『アーカイブズの構造認識と編成記述』思文閣出版、2014年、71-96頁。

(なかむら さとる 東京大学大学院新領域創成科学研究科)
 (ひえかた かずお 東京大学大学院新領域創成科学研究科)
 (みつゆき たいが 東京大学大学院工学系研究科)
 (かとう さとし 東京大学文書館)
 (みやもと たかし 東京大学文書館)
 (たかしま ともこ 東京大学大学院情報学環)