

Linked Data を用いた歴史研究者の史料管理と活用を支援するシステムの開発

中村 覚^{1,a)} 大和 裕幸^{2,b)} 稗方 和夫^{1,c)} 満行 泰河^{1,d)} 鈴木 淳^{1,e)} 吉田 ますみ^{1,f)}

受付日 2017年5月9日, 採録日 2017年11月7日

概要: 近年, デジタルアーカイブの普及にともない, インターネットを介した歴史資料 (以下, 史料) へのアクセスが容易となっている. 一方, 研究者が研究対象とする史料のすべてがインターネット上で公開されているとは限らず, また実証的な歴史研究が現存するすべての関係史料を検討することを基本にするため, 実際に図書館や図書館に赴き, 史料の収集を行う例も多い. 本研究では史料収集や整理に多大な労力を要する歴史研究の支援を目的とし, 複数の研究者の共同作業による史料収集・整理プロセスの効率化, および異なる専門知識を有する研究者の協調的な史料分析を支援するシステムを開発する. また, 外交文書の送付先の決定過程に関する歴史研究を行い, 開発したシステムの有用性を評価する.

キーワード: 歴史研究, Linked Data, メタデータ, 外交史

Development of System to support Management and Utilization of Research Data for Historical Researchers with Linked Data

SATORU NAKAMURA^{1,a)} HIROYUKI YAMATO^{2,b)} KAZUO HIEKATA^{1,c)} TAIGA MITSUYUKI^{1,d)}
JUN SUZUKI^{1,e)} MASUMI YOSHIDA^{1,f)}

Received: May 9, 2017, Accepted: November 7, 2017

Abstract: Digital archive has become popular recently, and this facilitates easy access to historical materials by the Internet. On the other hand, it is still common for researchers to gather materials by going to libraries and archives. This is because empirical historical research requires to examine all materials related to research objectives. In this study, the system is developed in order to gather and manage materials efficiently and support collaborative research with several researchers. This system manages materials on the Internet, and enables several researchers to accumulate and analyze materials collaboratively, which takes a great deal of effort individually. The effectiveness of the proposed system is evaluated by an experiment with several researchers to analyze the process of deciding addresses of diplomatic documents.

Keywords: historical research, Linked Data, metadata, history of diplomacy

1. はじめに

博物館や図書館等では, それらが保有する歴史資料 (以下, 史料) をデジタル化し, インターネット上で一般に公開するデジタルアーカイブの構築が進められている. また近年では, 史料の公開にとどまらず, 史料やメタデータの再利用・再配布を支援する Linked Data の活用が注目されている [1]. Linked Data とは, Web 上の機械可読なデータ, または当該データを実現させる仕組みのことであり, RDF (Resource Description Framework) 等の Semantic Web の

¹ 東京大学
The University of Tokyo, Bunkyo, Tokyo 113-8658, Japan
² 海上・港湾・航空技術研究所
National Institute of Maritime, Port and Aviation Technology, Mitaka, Tokyo 181-0004, Japan
a) nakamura.satoru@mail.u-tokyo.ac.jp
b) yamato-h@umisoraken.com
c) hiekata@edu.k.u-tokyo.ac.jp
d) mitsuyuki@sys.t.u-tokyo.ac.jp
e) jsuzuki@l.u-tokyo.ac.jp
f) be.lyk.nausicaa.20@gmail.com

標準技術を用いて表現される [2]. Europeana は、欧州の文化施設が保有する文化資源に関する情報を RDF に変換し、所蔵施設を横断した文化資源へのアクセスを可能としている。また、データに対する機械的なアクセスを可能とする API を提供し、分野やシステムを超えたコンテンツやメタデータの共有・利用を促進している [3].

このデジタルアーカイブの活用例の 1 つとして、史料を分析し新たな知見の獲得を目指す歴史研究における活用があげられる。歴史研究者（以下、研究者）がデジタルアーカイブを利用することにより、史料の収集に要する労力の削減等が期待できる。また、史料がデジタル化されている利点を生かし、史料に対する注釈付与を支援する研究 [4] や、情報技術を用いた史料分析に関する研究 [5] 等が数多く行われている。

一方、研究者が研究目的に応じて必要とするすべての史料がデジタルアーカイブ上で公開されているとは限らず、依然として研究者が図書館や文書館に赴き、史料を撮影する等して史料収集を行っている例は多い。三浦ら [6] は研究者へのインタビュー等を通じて、歴史研究プロセスを整理している。彼は史料の収集、整理、分析の 3 つのプロセスに大別し、それぞれ「各地に分散した史料収集」「史料から研究目的に応じた情報抽出」および「全体を俯瞰をした知見の獲得」の困難性を含むと述べている。研究者の個人作業が基本となる歴史研究では、史料の収集と整理に多大な労力を要し、数量が限られた史料に対する質的な分析が中心となっている。

本研究では上述した課題に対し、多大な労力を要する史料収集および整理プロセスの効率化、および歴史研究者や情報技術者等の異なる専門知識を有する研究者による協調的な史料分析を支援するシステムを開発する。研究者が収集した史料をインターネット上で管理し、複数研究者による史料の蓄積や分析を可能とする。さらに外交文書の送付先の決定過程に関する歴史研究を行い、提案するシステムの有用性を評価する。

2. 関連研究

2.1 情報技術を用いた歴史研究支援

近年、Europeana に代表されるように、Linked Data を用いたデジタルアーカイブの利活用を目的とした研究が数多く行われている [7], [8]. 久山ら [9] 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た。後藤 [10] は人間文化研究機構が持つ複数データベースを Linked Data を用いて再構築し、人文社会系の大規模データベースからのデータ発見を支援するプロトタイプを開発している。

史料がデジタル化された利点を生かし、情報技術を用いた史料分析を支援する研究も数多く存在する [11], [12]. 山田 [5] は、フィールドノートに記載されたテキストから

のトピック抽出を行い、フィールドノートが対象としている場面の特徴を抽出する手法を提案している。馬場 [13] は古地図の比較研究における困難性の解決を目的とし、デジタルアーカイブを利用した比較研究を行い、新たな古地図研究の方法論を提案した。鈴木ら [14] は、国立歴史民俗博物館が所蔵する錦絵コレクションを対象に、撮影したデジタル史料を測色値画像に変換し、画像データベースを構築した。また構築したデータベースを用い、錦絵画像の色彩分析を行った。北本ら [15] は研究の推論過程で積み重ねる証拠や過程をエビデンスネットワークというモデルに結合し、その分析から事実を導くための方法論を提案している。さらに、関野 [16] は時間情報の可視化や解析を支援する Web プラットフォーム「Hutime」を開発している。

また、史料に対する注釈付与を支援する研究も数多く行われている [17], [18]. 佐藤ら [4] は文章に対する注釈付与の効率化を目的とし、既存の注釈に関するデータを学習し、注釈が付与されていない文章に対して注釈の候補を提示するシステムを開発した。Yamato ら [19] は Linked Data を活用し、インターネット上で公開されている史料に対して、利用者の目的に応じたメタデータの付与や分析を支援するシステムを開発した。

一方、これらの情報技術を活用した歴史研究支援に関する方法論は、史料やメタデータに関するデータベースが存在することを前提としており、研究者が個々に収集・管理する史料への適用は困難である。iCloud や Google Drive 等の一般的なクラウドストレージを史料の収集や管理に利用することもできるが、これらのサービスは利用者が格納した画像やデータに対して自由にメタデータを付与するような利用方法は意図されていない。また、図書館や博物館、文書館等で利用されている既存のソフトウェアとして、Web パブリッシングシステムである Omeka [20] や、DSpace [21] 等のリポジトリシステムがあげられる。これらのソフトウェアは様々な形態のデジタル資料の登録や、Dublin Core をはじめとする各種スキーマに応じたメタデータの蓄積を可能とする。さらに、REST API や Linked Data 拡張による外部システムからのデータ利用を支援する機能も提供している。

本研究では、上述した研究成果や既存ソフトウェアが提供する機能群をベースとし、歴史研究を構成する史料の収集、整理、分析プロセスを一貫して支援するシステムを提案し、研究者の多様な目的に応じた史料管理や活用を支援することを目的とする。

2.2 公文書を用いた日本外交史研究

本研究では、第一次世界大戦以降の日本外交の検討を事例とする。当該期（大正期）の日本外交史研究は、政治家の個人文書が多く利用される明治期と比べて、日本外務省に残された公文書である「外務省記録」（外交史料館所

蔵)の利用の比重が大きいといえる。外務省記録は、明治初年から第二次世界大戦終結までに作成された電信、報告書等の外務省史料が約4万のファイルに編纂されたものであり[22]、1つのファイルに400~500枚の史料が収められていることも多くその史料数が膨大である。外交史研究者は、外務省が外務省記録の中の重要史料を翻刻・編纂した『日本外交文書』や『日本外交年表並主要文書』等の刊行物や自身が外務省記録から発見した在外公館報告や外務省本省の指示、訓令の内容を検討することで、外交交渉の場面をつぶさに研究してきた。

しかし、前述したように当然ながら個人で行う研究では読み込むことのできる史料数には限界があり、また内容の分析に注力するあまり個別の史料批判(当該史料が作成された経緯や処理方法を史料に残された形跡から判断すること)を怠ってきたために、内容自体の評価が疑われるという問題も存在している[23]。近年はこうした量的検討の軽視を批判し、史料学的アプローチから日本外交を再検討する研究がさかんである[24]、[25]。史料学的アプローチとは、外交文書に限らず近代史料(特に公文書)の作成、処理、保存過程の分析から意思決定プロセスを明らかにする方法であり、史料学は歴史学の基礎を構成すると同時に独自の学問分野としても成立している[26]。しかし、なおも個人研究の限界から史料の検討範囲が限定されており、より多くの史料を対象とした研究成果をあげることが課題となっている。

3. 歴史研究プロセスの整理と要件定義

ここでは本研究で対象とする歴史研究プロセスについて述べ、歴史研究支援に向けた要件定義を行う。

3.1 歴史研究プロセスの整理

まず三浦ら[6]が整理した史料の収集、整理、分析のプロセスについて、筆者らが研究者に対して行ったヒアリング結果と合わせて、各プロセスの内容を詳述する。史料の収集プロセスでは、研究者が図書館や文書館に赴き、各地に分散した史料を収集する。史料の収集方法としては、デジタルカメラやスマートフォンで史料を撮影し、計算機等に保存している。史料の整理プロセスでは、収集した史料の表題や出典等の書誌情報に加え、史料内に記述された情報を表計算ソフトやメモ帳等を用いて管理する。そして史料の分析プロセスでは、整理した史料の情報に基づき、史料の並べ替えや比較を行い、仮説の立案やその検証を繰り返すことで新たな知見の獲得を行う。

具体的な研究者の史料情報の管理例を図1に示す。図中の左部に示すように、収集した史料を計算機上に保存し、フォルダの階層に基づいた史料の分類を行っている。また右部に示すように、調査から得た情報はMicrosoft EXCELを用いて管理している。この方法では、史料とメタデータ

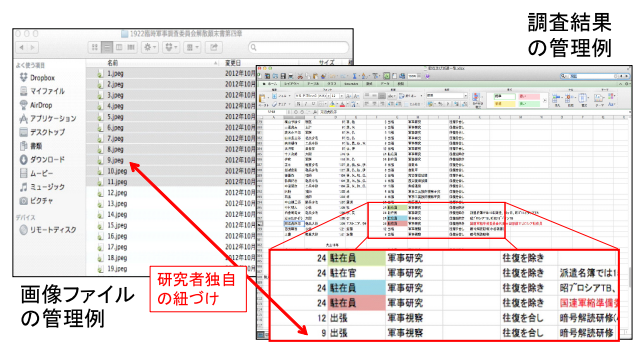


図1 研究者の史料と調査結果の管理方法例

Fig. 1 An example of methods to manage materials.

が異なる場所で管理されるため、史料を閲覧する際には当該史料を計算機上で検索する等の非効率的なプロセスが生じている。また各研究者の調査結果が研究者固有の環境に蓄積されるため、他の研究者との情報共有が困難となる。たとえば各研究者の調査結果や知見等は各人が独自に与えた識別子によって史料との紐づけが行われ、またそれらの記述形式は様々である。その結果、歴史研究は属人化し、各研究者の暗黙知に頼るがゆえの検索漏れや忘却による研究の低品質化、知識の風化等のリスクを含む。

3.2 歴史研究支援に向けた要件定義

上述した歴史研究に対し、複数の研究者の共同作業による史料の収集・管理プロセスの効率化、および異なる専門知識を有する研究者による協調的な分析支援を目的とし、以下に示す機能が必要となる。

- 複数の研究者が史料を共有可能な機能
- 史料の書誌情報等のメタデータを登録・管理する機能
- メタデータに基づく史料分類や並べ替え等を行う機能

1点目は複数研究者が史料収集プロセスを共同で進めるための機能である。研究者が個々に収集した史料を複数人で共有することで、各研究者の労力を低減すると同時に、研究対象とする史料数の増加を促す。2点目は史料の書誌情報や史料に記述された情報を整理するための機能である。史料は基本的に画像ファイルであり、検索や比較、分類を行うためには史料に関する情報、つまりメタデータをテキスト化して管理する必要がある。さらに複数の研究者による情報共有を行うにあたり、登録するメタデータの表記揺れを吸収する枠組みが必要となる。3点目は史料の分析を支援する目的で、蓄積したメタデータに基づく史料の分類や比較等を行うための機能が必要となる。

4. 提案手法

4.1 提案する歴史研究環境

3.2節で定義した要件に対し、本研究で提案する歴史研究環境を図2に示す。まず史料収集については、インターネット上にファイルサーバを構築し、研究者が収集した史

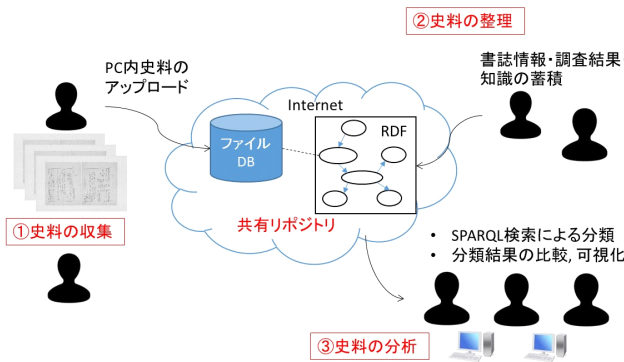


図 2 提案する歴史研究環境

Fig. 2 Proposed environment to support historical study.

料をアップロードすることで、インターネットを介した複数研究者による史料の共有と管理を可能とする。次に史料情報の整理については、史料の書誌情報や調査結果等のメタデータを RDF を用いて記述・管理する。RDF を用いる理由は、今後の普及が期待される Linked Data を用いたデータ公開を行うデータベースとの親和性が高いためである。また、研究者が対象とする史料や必要なメタデータ項目は研究目的に応じて多様であるため、ボトムアップ的なアプローチに対して表現方法の高い柔軟性を持つ RDF を採用した。さらに RDF スキーマを用いてメタデータの記述規約を与えることにより、複数研究者によるメタデータ登録時の表記揺れの吸収や情報共有の支援を行う。最後に RDF に対するクエリ言語である SPARQL を利用し、登録したメタデータに基づく比較や分類等の分析を可能とする。さらに、三浦が言及している「全体を俯瞰をした知見の獲得」を目的とし、分類結果の可視化を行い、研究者への気づきの提供や仮説の立案・検証を支援する。

4.2 史料情報の管理手法

ここでは RDF を用いた史料情報の管理手法について述べる。本手法では研究者が収集した史料を RDF モデルとして記述し、書誌情報や史料に記述された情報をそのメタデータとして管理する。図 3 に RDF モデルの例を示す。以下では図中に示す ① から ④ について順に説明する。なお語彙として、Dublin Core, FOAF に加え、DBpedia Japanese で用いられる接頭名詞空間「dbpedia-ja」、および本研究で独自に定義した接頭名詞空間「hm」を用いる。

① は研究者が収集・デジタル化した史料のファイルに関するメタデータを示す。各ファイルに対し、一意の URI を主語として持つ RDF モデルを対応付け、ファイル名や登録日等の情報をそのメタデータとして記述する。

② は研究過程で得た調査結果等、研究者が目的に応じて登録するメタデータを示す。各研究者が登録するメタデータに関して、それぞれ研究者ごとに Graph URI を与える。これにより、基本的には研究者が個々に研究成果を蓄積す

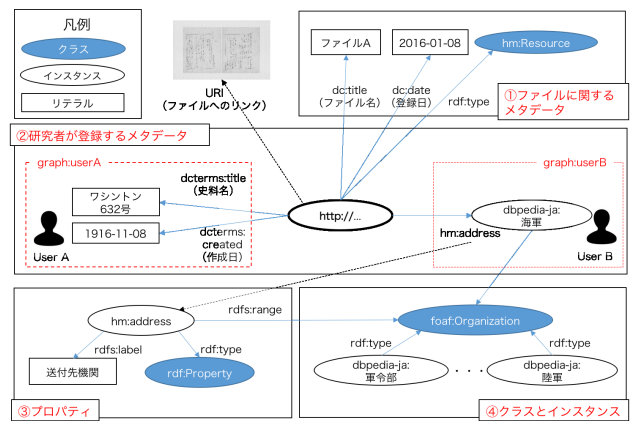


図 3 RDF を用いた史料情報の管理

Fig. 3 Management of information with RDF.

る。また、この Graph URI を研究グループごとに共有することにより、複数研究者間の情報共有を実現する。

研究者が登録するメタデータについては、プロパティと値のペアで登録する。その際、研究目的に応じて利用するメタデータの項目は異なるため、③ に示すように研究者が必要なプロパティを登録する。登録されたプロパティはすべて「rdf:Property」のインスタンスとして生成する。さらに必要に応じ、プロパティの値域の指定を行う。これにより④ に示すように、当該プロパティの値は値域に指定されたクラスのインスタンス群から選択する記述規約を与える。

なお本手法では、各研究グループに ID を与えて管理している。これは、研究グループごとに研究目的および利用するメタデータが異なることが想定され、また研究成果の新規性・独自性の観点から情報の共有範囲や公開時期を限定する必要があるためである。したがって、ID によって識別される各研究グループごとに、データやスキーマを独立して管理する。

5. 開発したシステム

開発したシステムの概要を図 4 に示す。本システムは Web アプリケーションであり、管理対象とするファイルのバイナリデータを格納するためのファイル DB と、4.2 節で述べた史料のメタデータを管理する RDF ストアの 2 種類の DB を提供する。また歴史研究に必要な史料の収集・整理・分析を支援する機能を提供する。以下ではこれらの各機能について説明する。

なお、本システムは Java を用いて開発し、RDF ストアとして Virtuoso[27]、RDF を取り扱うライブラリとして Jena[28] を利用した。

5.1 史料の収集

史料の収集を支援する機能としては、撮影やダウンロードによって得た史料をシステムに登録するファイルアップ

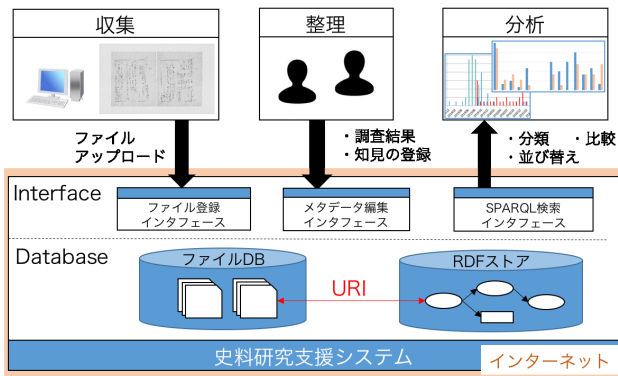


図 4 開発した歴史研究支援システム

Fig. 4 Developed system to support historical study.



図 5 メタデータ編集インタフェース

Fig. 5 Interface for editing metadata of materials.

ロード機能を提供する。アップロードされたファイルは、システム内部で図 3 ①部に示した RDF モデルに変換される。登録時に各ファイルに一意の URI が与えられ、ファイル名や登録日等の情報がメタデータとして自動的に付与される。さらにファイルのバイナリデータはファイル DB に格納され、付与された URI を参照することで、当該ファイルへアクセスすることができる。

5.2 史料の整理

史料の整理を支援する機能としては、図 5 に示すメタデータ編集インタフェースを提供する。本インタフェースは画像とメタデータを同一画面で表示することで、ユーザは史料の閲覧と並行してメタデータを編集することができる。画像の閲覧画面では、画像の拡大・縮小・回転等が可能である。メタデータ編集機能では、登録済みのメタデータの更新や削除を行うための機能と、新たにメタデータを登録するための入力フォームを提供する。

また値域を指定したプロパティに関しては、当該クラスのインスタンス群から選択して値を登録する。これらのプロパティ、クラス、インスタンスについては、本システムが別途提供する RDF スキーマの管理インタフェースを用いて、新規データの登録や既存データの編集を行う。たとえば図 5 中の「送付先機関」というプロパティについては、

値域として「組織」に関するクラスが指定されており、「組織」クラスのインスタンスである「海軍」、「軍令部」、「陸軍」等から選択して登録する。このスキーマ定義については、研究グループ内のすべてのユーザが実行する権限を持つ。基本的には、研究グループ内の参加者がコミュニケーションを取り、研究目的に応じて必要となるプロパティや値域の定義を行う。また、史料の整理プロセスを通じ、新たな知見や分析観点の変化によって新たな項目が必要となる度に、スキーマの追加や修正を行う。

5.3 史料の分析

史料の分析については、SPARQL Endpoint を利用し、目的に応じた史料のメタデータを抽出する。たとえば、結果として得られたメタデータの一覧を Microsoft EXCEL 形式でダウンロードし、グラフ化による可視化や比較、並べ替え等による史料分析を行う。また、歴史研究者と情報技術者のように、異なる専門を有する参加者によって研究グループが形成されている場合には、情報技術者が史料分析の支援を目的としたアプリケーションを構築する。主として研究者が蓄積するデータに対して、SPARQL Endpoint を用いた柔軟なアクセスが可能となることにより、技術者による多様な研究目的に応じた分析手法の実装を支援する。次章の中で、この分析方法の具体例を述べる。

6. 実験：外交文書の送付先決定過程の分析

ここでは、歴史研究プロセスに対するシステムの有用性を評価するにあたり、開発したシステム上で実施した実験内容について述べる。

6.1 実験背景

膨大な研究蓄積を有する戦間期日本政治外交史研究の大きな 2 つの潮流として、内閣総理大臣や外務大臣、外務次官といった個人の外交理念や政治能力に日本の外交政策のあり方を規定させるもの [29], [30] と、軍部あるいは政軍関係から対外政策の帰趨を描くもの [31], [32] がある。サブプレーヤとして描かれる外務省は、政策立案機能や所謂「革新派」外交官の政策構想がクローズアップされそれ以外の役割や機能に着目されてこなかったため、政府内の政治構造における位置づけは解明されないまま「軍部の圧力に屈して外交を統制できなかった外務省」像が固定化されている。しかし「軍部の圧力」を日本外交の失敗要因として絶対視している限り明治憲法体制という前提のもとでの日本外交の根本的、構造的問題は明らかにできないのであり、日本政治外交史は膠着状態にあるといえる。

前述したように従来の外交史研究においては『日本外交文書』に翻刻掲載された電信史料が多く参照されてきたが、電信の内容のみを翻刻した同書に依拠した研究手法では省外への情報送付の書き込みの存在を認識できず、来電の省

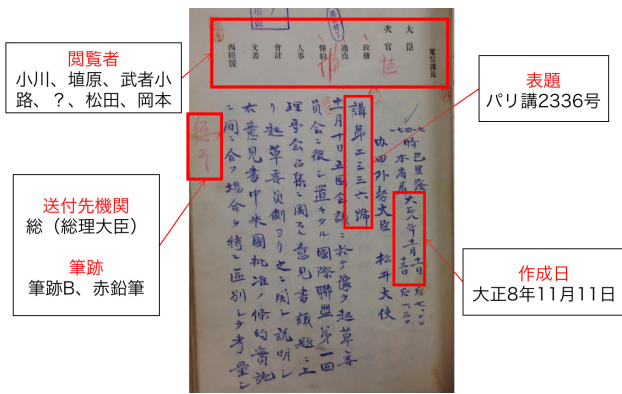


図 6 分析対象とした来電の例

Fig. 6 Example of diplomatic cable.

内での処理方法と省外への提供の実態を再現することができない。本研究では、外務省記録の中から 1918 年前後～1923 年を対象として、第一次世界大戦のパリ講話会議の動向報告およびワシントン会議開催前から開催中にかけてのアメリカ政府内の動向報告と、電信 1 枚ずつにメモ書きされた省外への送付指示を分析対象とし、筆跡や筆記具、閲覧者を手がかりとする史料学的アプローチをとることにより、外務省と省外組織の関係性の一端を明らかにしていく。

6.2 実験内容

以上の前提をふまえ、本実験では第一次世界大戦後の来電に関する送付先決定過程の分析を行う。来電とは在外の大使、公使等の駐在員から外務本省へ届いた電信であり、外務省内で閲覧された後、総理大臣・海軍・陸軍等の他機関への送付処理がなされた [33]。外務省はこれらの情報の送付先決定権を有していたが、官僚制の運用の中で実際に送付先を決定していた人物・役職については史料上からは明らかになっていない。この送付先決定者を特定することを目的とし、来電に記述された閲覧者や筆跡に関する情報をメタデータとして整理し、それらを用いて史料の分類・比較を行う。なお、本実験は歴史研究者（以下、研究者 X）、情報技術者（以下、技術者 Y）、および史料収集の補助要員 2 名の計 4 名で行った。

6.3 史料の収集

史料の収集に関して、外交史料館が保有しているパリ講和会議に関する来電を収録した簿冊 10 件、計 471 件の来電を収集した。具体的には、外交史料館にて技術者 Y、および史料収集補助要員 2 名が対象史料を撮影し、5.1 節のファイルアップロード機能を用いて画像ファイルをシステムに登録した。登録した来電の例を図 6 に示す。各来電には史料名と作成日のほか、史料上部にはその来電を閲覧した人物のサイン、史料左部にはその来電が送付された機関の名称が記述されていることが確認できる。

表 1 来電に関する情報記述のためのスキーマ

Table 1 Defined schema to describe information of materials.

プロパティ名	ラベル名	値域
dcterms:title	表題	
dcterms:created	作成日	
hm:reader	閲覧者	foaf:Person
hm:address	送付先機関	foaf:Organization
hm:handwritingType	筆跡	

クラス名	ラベル名	インスタンス例
foaf:Person	人物クラス	dbpedia-ja:埴原正直
foaf:Organization	組織クラス	dbpedia-ja:海軍 dbpedia-ja:陸軍

6.4 史料の整理

史料の整理に関しては、主に研究者 X がシステムにアクセスし、6 章に示した来電に記述された情報をメタデータとしてシステムに登録した。メタデータの登録にあたり、定義したプロパティやクラスの一部を表 1 に示す。

まず、研究者 X が史料の読解プロセスを通じ、収集した来電から確認可能な項目について、システムにプロパティに登録した。具体的には、来電の表題と作成日を記述するためのプロパティとして「dcterms:title」、「dcterms:created」をそれぞれ与えた。また史料の閲覧者、送付先機関、送付先の筆跡に関しては、接頭名詞空間「hm」を独自に定義し、それぞれ「hm:reader」、「hm:address」、「hm:handwritingType」を定義した。

加えて、前者 2 つのプロパティに関しては値域を定義し、それぞれ人物クラス「foaf:Person」、組織クラス「foaf:Organization」を指定した。たとえば組織クラス「foaf:Organization」のインスタンスとして、「dbpedia-ja:海軍」や「dbpedia-ja:陸軍」等を登録し、メタデータの値として利用した。このような値域の定義を行った理由は、本プロセスと並行して技術者 Y が担当していた分析プロセスにおいて、閲覧者や送付先機関に関する値の構造化が求められたためである。たとえば実験開始当初、「送付先機関」の値として「総陸海（総理大臣・陸軍・海軍の頭文字）」が登録され、機関別の集計等に適さないことが分かった。したがって、研究者 X と技術者 Y の間で記述形式に関する検討を行い、上述したプロパティに関して値域を定義した。これにより、以後研究者 X を主とするデータ登録者が図 5 に示したメタデータ編集インタフェースにおいてプロパティを選択した際、値域に指定されたクラスのインスタンス群から選択して登録することで、値の表記揺れを吸収した。

6.5 史料の分析

6.4 節における史料整理の結果、送付先指示筆跡として A, B, S の 3 種類が確認された。これらの各筆跡の人物を

史料に関するURI	作成日	筆跡
uri	created	handwritingType
http://172.16.53.71:48080/diplomatic/resource/get?resourceID=1129	20131217	A, 筆
http://172.16.53.71:48080/diplomatic/resource/get?resourceID=1130	20131217	A, 筆
20131217	20131217	A, 筆
20131217	20131217	B, 鉛筆
20131217	20131217	A, 筆
20131217	20131217	A, 筆
20131217	20131217	A, 筆
20131204	20131204	筆跡S
http://172.16.53.71:48080/diplomatic/resource/get?resourceID=1178	20131217	A, 筆
http://172.16.53.71:48080/diplomatic/resource/get?resourceID=118	20131204	筆跡S
http://172.16.53.71:48080/diplomatic/resource/get?resourceID=1181	20131217	A, 筆


```

SELECT DISTINCT ?uri ?created ?handwritingType
WHERE {
    ?uri dct:created ?created .
    ?uri hm:handwritingType ?handwritingType .
}
    
```

SPARQLクエリ

図 7 SPARQL による筆跡と作成日の抽出

Fig. 7 Extraction of handwriting types and creation date by SPARQL.

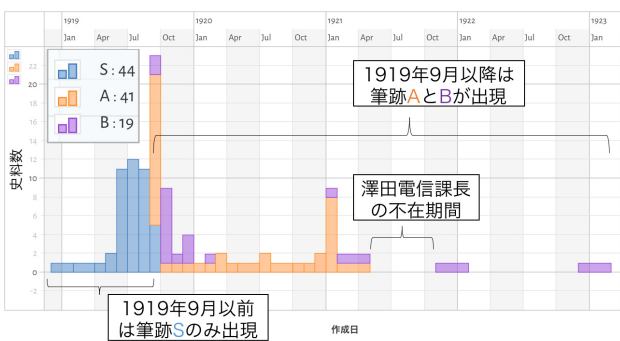


図 8 筆跡と作成日の関係性の可視化

Fig. 8 Analysis based on the relation of handwriting types and creation date.

特定するため、筆跡と作成日の関係、および筆跡と閲覧者の関係の2つの観点から分析を行った。なお、これらの分析観点の設定においては、研究者 X と技術者 Y がコミュニケーションを取りながら検討し、技術者 Y が SPARQL クエリの記述、および抽出結果の可視化を行うアプリケーションを構築することによって実装した。

6.5.1 筆跡と作成日の関係に基づいた分析

まず来電を時系列に並べ替えることで、筆跡の出現時期に基づく分析を行った。5.3 節で説明した分析機能を用い、技術者 Y が図 7 に示す SPARQL クエリを発行し、来電の作成日と送付先指示筆跡に関するメタデータを抽出した。さらに、その抽出結果を JSON ファイルとして取得し、ブラウザ上で可視化するアプリケーションを作成した。その出力結果を図 8 に示す。

この結果から筆跡 S は 1919 年 9 月以前にのみ登場していることが確認できる。これは幣原喜重郎（以下、幣原）が外務次官を務めていた時期であり、この期間では外務次官が送付先を決定していたという仮説を得ることができた。一方、埴原正直（以下、埴原）が外務次官に就任した 9 月以降では、少なくとも筆跡 A, B の 2 種類の筆跡が出現しており、外務次官以外の人物が来電の送付先指示に関わっていた可能性が示唆される。

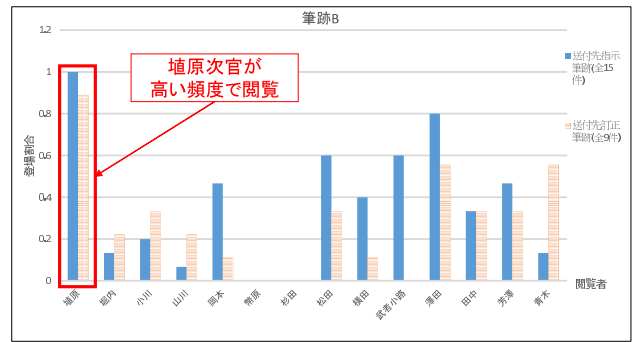


図 9 筆跡別の来電における閲覧者の割合

Fig. 9 Frequency of readers in cables of each handwriting type.

以下では筆跡 A, B の人物特定を目的とし、それぞれの筆跡で送付先機関が指定されている史料に対する分析を行う。

6.5.2 筆跡と閲覧者の関係に基づいた分析

次に筆跡別の来電における閲覧者の登場割合に基づく分析を行った。これは研究者 X が史料の整理プロセスにおいて、埴原が来電閲覧時に記入するサインと送付先を記述した筆跡 B が類似すると気づいたためである。この分析結果を図 9 に示す。このグラフについては、6.5.1 項と同様に技術者 Y が送付先に関する筆跡別の閲覧者を抽出し、Microsoft EXCEL を用いて可視化したものである。この結果から、筆跡 B によって送付先が指示（15 件）または訂正（9 件）されたほぼすべての来電において、埴原が閲覧していることが分かる。これは研究者 X が立てた筆跡 B が埴原であるという仮説を支持する 1 つの量的な結果である。

この発見に基づき、筆跡 A による送付先の指示が行われている来電に対しても同様の分析を行った。この結果、当時電信課長であった人物「澤田」が高い割合で閲覧していることが確認でき、筆跡 A は澤田であるという仮説を立てた。この仮説の検証を目的とし、研究者 X が当時の澤田の動向を他文献 [34] を用いて調査した結果、1921 年 3 月 3 日から 9 月 3 日まで皇太子の外遊に同行していることが判明した。この期間に出現した筆跡を図 8 で確認した結果、当該期間には筆跡 A による送付先が指示された来電が存在しないことが確認でき、筆跡 A が澤田のものであるという仮説を支持する結果を得ることができた。

6.6 結果

以上の分析結果から、筆跡 A は澤田電信課長、筆跡 B は埴原次官、筆跡 S は幣原次官であるという仮説を立案することができた。またこれらの発見に基づき、幣原次官期には外務次官が来電の送付先を単独決定していたのに対し、埴原次官期には外務大臣への悉皆的回覧が廃止され、送付先となる外部機関の選定権限は外務次官から電信課長へ委譲された、という来電の送付先決定過程の変化に関する仮

表 2 筆跡別の来電送付先機関の割合

Table 2 Addresses in cables of each handwriting type.

送付先決定者	幣原次官 (筆跡 S)	埴原次官 (筆跡 B)	澤田電信課長 (筆跡 A)
総来電数 (件)	44	19	41
送付先機関 (件)			
総理大臣	44	18	41
参謀本部	43	8	27
海軍省	43	9	24
軍令部	43	9	27
陸軍省	43	9	24

説を立案した。このような変化が生じた理由としては、電
信量の増加にともなう事務の合理化の結果と考えられる。

さらに、送付先決定者ごとの送付先機関の割合を表 2 に
示す。この結果から、今回対象とした史料においては幣原
次官期には内容を問わずすべての政府内組織へ平等に電信
情報が送付されていたのに対し、埴原次官期には送付先
の取捨選択が行われていることが分かる。その判断基準は情
報の表面的内容を基本とする属人的なもので、また全体と
して首相（原敬）限りの送付が増加する傾向にあった。

すなわち、大戦後の省内の事務合理化の結果、外務省に
よる省外への情報提供において、提供内容の選択に次官以
下の外務官僚の恣意性、主体性が反映されるようになった
ことが明らかになった。原首相は外務省の政治的後見者で
あったとされる [35] が、電信の原首相への集中は外務省
から原へのベクトルを示すものとして相互の近接性を実証す
るものである。熊本 [36] は省外への電信送付分析から幣原
次官期に政策方針の省内不統一が省外への情報提供の停止
を招いたことを論じたが、送付判断は次官が行ったものと
暗に仮定している。本研究の成果はこの仮定を裏付けると
同時に幣原以後の次官人事の変更が情報提供のあり方をめ
ぐる判断主体（役職）の変更を招いたことを示すものであ
り、当該期の外務省の情報管理はその内容だけでなく人事
や政治的關係にも左右され、制度として至極流動的であっ
たことを指摘しうるものである。

7. 考察

7.1 Linked Data の利点

本研究では、Linked Data を用いた資料管理手法を提案
した。ここでは、Linked Data の仕組みが、史料の整理や
分析等の歴史研究プロセスに与える利点について考察する。

7.1.1 史料の整理

歴史研究の特徴の 1 つとして、研究目的の多様性があげ
られる。具体的には、史料の収集や読解等の整理プロセス
における新たな知見や分析観点の変化に応じたスキーマ更
新等のボトムアップ的なアプローチが求められる。これに
ついて、RDF のデータ記述の柔軟性を活用することによっ
て、研究の進行に合わせてプロパティの追加や更新が容易

となった点が、Linked Data を活用した利点の 1 つである。

また、複数人が並行してメタデータの登録プロセスを遂
行するにあたり、メタデータのプロパティに関して値域の
指定等の記述規約を与えることにより、背景知識の異なる
複数研究者が登録するメタデータの表記揺れを吸収した。
実験においては、情報技術者が史料の比較や分類に適した
記述形式をスキーマとして定め、歴史研究者がその記述規
約に従って情報入力を行うことにより、研究者が登録した
情報に基づく量的な史料分析が可能となった。これらの観
点から、登録するメタデータに関する記述規約を与える管
理手法は有効に機能した。ただし、すべてのプロパティの
対して値域を定義することは、自由な記述による分析可能
性を阻害するリスクを持つ。そのため、歴史研究者による
自由な値の登録を基本とし、定量的な分析に使用する項目
のみ値域を定義することにより、研究の多様性を確保しつ
つ、効率的な史料分析を支援する。

一方、本研究では、Linked Data の仕組みを通じた外部
データとの連携や活用には至らなかった。しかし、歴史史
料がアジア歴史資料センター、国立公文書館、国立国会図
書館、そして各府県の図書館や大学等でそれぞれ異なった
方式でデジタル公開されている現状、および実証的な歴史
研究が現存するすべての関係史料を検討することを基本に
するため、それら複数の異種データベースと未デジタル化
の史料を並列的に参照する必要がある。この観点から、外
部データとの連携が意図された Linked Data による史料情
報の管理は、今後より重要性が増していくと考えられる。

7.1.2 史料の分析

史料の分析プロセスに関しては、研究者が登録したメタ
データに基づいた分類を行い、分類結果の可視化等に基づ
く分析を可能とした。この結果、歴史研究者と情報技術者
という異なる分野の研究者による協調的な研究環境が実現
した。具体的には、6 章の実験において、研究者が立案し
た仮説に対する情報技術者による量的な検証、および情報
技術者による分析結果の可視化を通じた研究者への仮説
の提供を実現した。この観点から従来の研究者個人で行わ
れていた分析と比較し、本研究で提案した歴史研究手法の
有用性が示唆された。特に、SPARQL Endpoint の利用に
よって、研究者の多様な観点や目的に応じて、情報技術者
によるデータ抽出や分析結果の可視化等を行うアプリケー
ションの構築が容易となった。この点は、システム外部か
らのデータ活用を支援する Linked Data を利用した利点で
ある。

一方、史料分析の目的や方法の多様性を考慮し、本シ
ステムは利用者に SPARQL クエリの記述を求める方式を
とった。本実験では歴史研究者と情報技術者が共同で実験
を行ったため問題は生じなかったが、研究者をはじめとす
る利用者が SPARQL クエリを記述することは一般的に困
難である。今後は SPARQL 等の技術に馴染みのない利用

者を対象とした分析支援インタフェースを開発し、研究者が単独で利用可能なシステムの構築を目指す。

7.2 開発したシステムに対する考察

本実験では研究者が分析対象とする史料をインターネット上のデータベースで管理し、史料収集プロセスを複数人が並行して遂行可能な環境を構築した。実験では、簿冊1冊あたり90分程度の時間を要する史料収集プロセスを3名で分担することで、10冊の簿冊から500件弱の史料を3日間で収集した。これにより、作業分担による研究者の史料収集における労力の軽減に加え、比較的多量の史料に基づく量的な分析が可能となった。

史料の整理プロセスに関しては、特に史料のコンテンツと書誌情報を同一画面に表示し、史料の閲覧と並行したメタデータの登録や編集を可能とするインタフェースについて、実験に参加した研究者から高く評価された。この理由として、図1で述べた従来の画像コンテンツと調査結果を異なる環境で管理する方法に比べ、登録したメタデータに基づく史料の検索や管理が容易となった点があげられた。

7.3 史料の媒体や構成の違いによる提案手法の適用範囲

本研究では、第一次世界大戦後の外交文書（公文書）を対象とした実験を行い、提案手法の評価を行った。具体的には、同様の記述形式を持ち、かつ同時期に作成された史料を研究対象とした。本手法は、史料の形式的な情報に基づく史料の比較や分類を支援することを目的としている。この観点において、媒体や構成が異なる複数の史料（たとえば、文書、古記録、絵巻等）を対象とした研究に対しても適用可能である。Yamatoら[19]は海軍造船官であった平賀譲が遺した史料群を対象とした研究を行い、設計図面や技術計算書等の媒体が異なる史料群に対する分析を行っている。ただし、史料の記述内容を比較・検討する場合には、各々の媒体や構成に応じて異なる意味論を持ちうる。この場合には、量的に比較可能な情報だけでなく、「翻刻文」や「研究メモ」といったメタデータによって質的な情報を管理し、量・質の両観点に基づく史料分析へ応用する。

7.4 オープンサイエンスへの活用

近年オープンサイエンスという概念の普及にともない、歴史研究を含む人文科学の分野においてもデータのオープン化が進められている[37], [38]。データのオープン化は、データの共有による研究促進や、研究の反証可能性を高めることによる研究成果の信頼性の向上等への寄与が期待されている。このような動向において、本研究では研究者が収集した史料や目的に応じて整理したデータをLinked Dataを用いて管理し、データに基づく仮説の立案と検証を実現した。また、本実験で利用したデータはSPARQL Endpointを用いて再利用可能であり、さらに取得したデー

タを用いた実験結果の再現が可能である。公開方法（範囲や時期）およびデータの著作権等解決すべき課題は多いが、本研究が歴史研究におけるオープンデータの活用基盤の1つとして位置付けられると考える。今後は本手法を応用し、人文科学におけるオープンサイエンスの実践に取り組む。

8. 結論

本研究では複数研究者が史料の収集・整理・分析を共同で遂行し、歴史研究プロセスの効率化および異なる専門知識を有する研究者による協調的な史料分析を可能とするシステムを開発した。研究者が収集した史料をインターネット上で管理し、複数研究者による史料の蓄積や分析が可能な環境を構築した。さらに外交文書の送付先の決定過程に関する歴史研究を通じ、開発したシステムが史料収集プロセスの労力低減、RDFを用いた史料情報の管理と記述規約に基づく情報共有支援、および量的な史料分析による仮説の立案と検証支援に寄与することを確認した。今後は異なる研究目的を持つグループが混在するような研究者群を対象とした評価実験を行い、提案手法の有用性およびシステムの改良を検討する。

参考文献

- [1] 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, Vol.58, No.2, pp.127-134 (2015).
- [2] Berners-Lee, T.: Linked Data, available from (<https://www.w3.org/DesignIssues/LinkedData.html>) (accessed 2017-04-22).
- [3] Europeana Creative Challenges. available from (<http://pro.europeana.eu/get-involved/projects/project-list/europeana-creative/challenges>) (accessed 2017-04-22).
- [4] 佐藤貴文, 後藤 真, 木村文則, 前田 亮：歴史資料からの人文系研究者への注釈候補の提示手法の構築, じんもんこん 2015 論文集, Vol.2015, pp.165-172 (2015).
- [5] 山田太造：フィールドノートに記述された場面を特徴付ける—語彙による知識処理, 情報知識学会誌, Vol.25, No.4, pp.315-324 (2015).
- [6] 三浦 崇, 伊東幸宏, 小西達裕, 田村貞雄, 赤石美奈, 中谷広正, 阿部圭一：歴史学研究支援システムの構築, 情報処理学会研究報告, Vol.97, No.48, pp.25-30 (1997).
- [7] 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, Vol.2012, No.7, pp.239-246 (2012).
- [8] 研谷紀夫, 馬場 章：文化資源オントロジーの構築とその活用 (〈特集〉第15回 (2007年度) 年次大会 (研究報告会&総会)), 情報知識学会誌, Vol.17, No.2, pp.129-134 (2007).
- [9] 久山岳夫, Biligsaikhan, B., 木村文則, 前田 亮：浮世絵を対象とした異種データベースの多言語統合アクセス手法の提案, じんもんこん 2011 論文集, Vol.2011, No.8, pp.275-280 (2011).
- [10] 後藤 真：人文社会系大規模データベースへのLinked Dataの適用—推論による知識処理, 情報知識学会誌, Vol.25, No.4, pp.291-298 (2015).

- [11] 倉持 基, 研谷紀夫, 津田光弘, 馬場 章: デジタルアーカイブを利用した歴史写真の情報学的研究, *じんもんこん 2005 論文集*, Vol.2005, pp.113-120 (2005).
- [12] Yu, C.-H. and Hunter, J.: Documenting and Sharing Comparative Analyses of 3D Digital Museum Artifacts Through Semantic Web Annotations, *J. Comput. Cult. Herit.*, Vol.6, No.4, pp.18:1-18:20 (2013).
- [13] 馬場 章: デジタルアーカイブを利用した地図の書誌学—長久保赤水製作『改正日本輿地路程全図』の場合, *じんもんこん 2003 論文集*, Vol.2003, pp.119-126 (2003).
- [14] 鈴木卓治, 安達文夫, 大久保純, 小林光夫: 錦絵資料の測色値画像データベースの構築と色彩分析の試み, *じんもんこん 2004 論文集*, Vol.2004, pp.75-82 (2004).
- [15] 北本朝展, 西村陽子: Digital Criticism Platform: エビデンススペースの解釈を支援するデジタル史料批判プラットフォーム, *じんもんこん 2015 論文集*, Vol.2015, pp.211-218 (2015).
- [16] 関野 樹: Web HuTime—時間情報のための Web プラットフォーム, *じんもんこん 2016 論文集*, Vol.2016, pp.125-132 (2016).
- [17] Nagasaki, K., Tomabechi, T. and Shimoda, M.: Towards a digital research environment for buddhist studies, *Literary and Linguistic Computing*, Vol.28, No.2, pp.296-300 (2013).
- [18] Donato, F.Di., Morbidoni, C. and Fonda, S.: Semantic annotation with Pundit: A case study and a practical demonstration, *Proc. DH-CASE'13*, ACM (2013).
- [19] Yamato, H., Hiekata, K., Mitsuyuki, T. and Nakamura, S.: Historical Design Review based on the Digital Archive with the Semantic Web Approach, *Proc. 11th International Marine Design Conference*, Vol.1, pp.457-464 (2012).
- [20] Omeka: available from <http://omeka.org/> (accessed 2017-08-09).
- [21] DSpace: available from <http://dspace.org/> (accessed 2017-08-09).
- [22] 戦前期「外務省記録」: 入手先 <http://www.mofa.go.jp/mofaj/annai/honsho/shiryo/shozo/index.html> (参照 2017-08-09).
- [23] 小池聖一: 近代日本文書学研究序説, p.107, 現代史料出版社 (2008).
- [24] 熊本史雄: 大戦間期の対中国文化交流, 吉川弘文館 (2013).
- [25] 佐藤元英, 服部龍二 (編著): 日本外交のアーカイブ学的研究 2, 中央大学出版部 (2016).
- [26] 中野目徹: 近代史科学の射程, pp.11-13, 弘文堂 (2000).
- [27] OpenLink Virtuoso, available from <https://github.com/openlink/virtuoso-opensource> (accessed 2017-04-22).
- [28] Apache Jena, available from <https://jena.apache.org/> (accessed 2017-04-22).
- [29] I. ニッシュ: 日本の外交政策, ミネルヴァ書房 (1994).
- [30] 服部龍二: 東アジア国際環境の変動と日本外交, 有斐閣 (2001).
- [31] 北岡伸一: 日本陸軍と大陸政策, 東京大学出版会 (1978).
- [32] 小林道彦: 政党内閣の崩壊と満州事変, ミネルヴァ書房 (2010).
- [33] 千葉 功: 旧外交の形成, 勁草書房 (2008).
- [34] 外務省: 外務省年鑑: 大正 11 年, p.292, クレス出版 (1999).
- [35] 三谷太一郎: 日本政党政治の形成, 東京大学出版会 (1967).
- [36] 熊本史雄: 大戦間期外務省の情報管理と意思決定, *日本史研究*, No.653, pp.3-37 (2017).
- [37] 北本朝展, 山本和明: 人文学データのオープン化を開拓する超学際的データプラットフォームの構築, *じんもんこん 2016 論文集*, Vol.2016, pp.117-124 (2016).
- [38] 嘉村哲郎, 大向一輝: 人文科学における Linked Open Data

の活用, 人工知能, Vol.32, No.3, pp.401-407 (2017).



中村 覚 (正会員)

2017年東京大学大学院新領域創成科学研究科博士後期課程修了。博士(環境学)。同年東京大学情報基盤センター学術情報研究部門助教。現在に至る。Linked Dataをはじめとする情報技術を用いたデジタルアーカイブの構築と活用に関する研究を行っている。Alliance of Digital Humanities, 日本船舶海洋工学会等の会員。



大和 裕幸

1954年5月14日生。1982年東京大学大学院工学系研究科船舶工学専門課程博士課程修了, 工学博士。同年科学技術庁航空宇宙技術研究所(当時)研究員, 短距離離着陸実験機「飛鳥」の研究開発に従事。88年東京大学工学部船舶工学科助教授, 97年船舶海洋工学専攻教授, 09年新領域創成科学研究科長, 13年理事・副学長を経て, 16年より国立研究開発法人海上・港湾・航空技術研究所理事長。東京大学名誉教授。日本船舶海洋工学会会長, 日本工学会理事, 日本学術会議連携会員等を歴任。



稗方 和夫 (正会員)

1974年12月17日生。2000年東京大学大学院工学系研究科環境海洋工学専攻修士課程修了。同年4月より日本アイ・ビー・エム株式会社入社, ソフトウェア開発研究所にてLinux OSの開発に従事。2004年東京大学大学院工学系研究科環境海洋工学専攻助手。2007年4月同助教等を経て2010年2月より東京大学大学院新領域創成科学研究科准教授。2006年(独)情報処理推進機構よりスーパークリエイター認定。2007年日本船舶海洋工学会奨励賞(乾賞)受賞, 2014年日本船舶海洋工学会論文賞受賞, 所属学会は, 日本船舶海洋工学会, 人工知能学会, 日本機械学会, Society of Naval Architects and Marine Engineers 等。



満行 泰河 (正会員)

2014年東京大学大学院新領域創成科学研究科博士後期課程修了。博士(環境学)。2015年より東京大学大学院工学系研究科助教。現在に至る。設計工学, システム工学に関する研究活動を行っている。日本船舶海洋工学会, 日本機械学会等の会員。



鈴木 淳

1962年生。1992年東京大学大学院人文科学研究科博士課程単位取得済退学。1995年文学博士。1992年東京大学社会科学研究所助手, 1994年東京大学教養学部助教授, 1999年東京大学大学院人文社会系研究科助教授, 2007年准教授, 2013年教授。所属学会は, 史学会, 日本産業技術史学会, 政治経済学・経済史学会等。



吉田 ますみ

東京大学大学院人文社会系研究科博士課程在籍中。日本学術振興会特別研究員(DC1)。主に1920~1930年代の日本政治外交史を専門としている。史学会, 日本国際政治学会各会員。