

# TEI と IIF をベースとしたオン/オフライン併合型史料研究支援システムの開発 - オスマン・トルコ語文書群を対象として

中村 寛 (東京大学情報基盤センター) ・ 佐治 奈通子 (東京大学大学院人文社会系研究科)

永崎 研宣 (人文情報学研究所)

本研究では、TEI と IIF を用いて歴史学研究者の史料整理と分析を支援するためのシステムを開発した。さらに既存の関連ツール群を最大限利用すること、および研究者の作業環境を考慮し、オンラインとオフラインを併合した研究環境を提案した。TEI を用いた史料情報の記述、および可視化による史料の整理・分析支援を行うことで、歴史学研究者に対して新たな気づきを提示し得ることを確認した。

## Development of system to support historical study with on-offline combined tools based on TEI and IIF - A case of Ottoman and Turkish documents

Satoru Nakamura (Information Technology Center, The University of Tokyo)

Natsuko Saji (Faculty of Humanities, The University of Tokyo)

Kiyonori Nagasaki (International Institute for Digital Humanities)

In this study, we developed a system to support historical study using TEI and IIF. In addition, we proposed an on-offline combined research environment, taking into account the maximum use of existing related tools and the working environment of researchers. It was confirmed that the developed system helps researchers to bring new awareness by describing the contents of historical documents with TEI and supporting the analysis of those documents with visualization.

### 1. はじめに

筆者らは史料研究プロセスにおける史料管理および分析を支援するシステムの開発に取り組んできた[1]。これまでは Linked Data を用いた史料のメタデータ管理に主眼を置いてきたが、本研究では史料の本文記述の整理および分析への活用支援を目的とする。具体的には、IIF を用いて史料の画像データを管理し、TEI を用いて史料本文を記述する。史料本文の作成、管理を支援するサービスとして、みんなで翻刻[2]や FromThePage[3]、Recogito[4]などがある。例えば Recogito では、プリセットのタグが提供されているなど、TEI に馴染みのない利用者にとっても利用しやすいものとなっている。一方、想定される利用者・利用方法に対応すべく使用可能な TEI 要素が限定されている。本研究では TEI に馴染みのある利用者を対象として、TEI を直接編集することにより、史料情報を記述・管理する研究環境を提案する。また、先に述べたサービスは史料本文の作成に主眼を置いているが、本研究では、史料本文から抽出した構造化データを史料分析へ利用することを前提としたシステムを開発する。

さらに本研究は、既存の関連ツール群を最大限利用すること、および研究者の作業環境を考慮し、オンラインとオフラインを併合した研究環境を

提案する。具体的には、IIF 画像の発信やメタデータの管理、各種史料情報の可視化環境についてはサーバ上に構築し、オンラインでの利用を前提とする。この目的として、Omeka[5]等の既存ツールが提供する IIF 画像サーバ機能やメタデータ検索・管理機能が利用できる、研究者毎の環境構築が不要となる、研究者間のデータ共有が容易となる、等が挙げられる。一方、TEI を用いた史料本文の記述プロセスについては、デファクトスタンダードツールである Oxygen[6]を用いたオフライン環境を利用する。この利点として、既存の TEI 編集用オンラインツールと比較して、Oxygen が提供する多様な機能が使用できることに加え、利用者がインターネットに接続できない環境においても史料本文の作成が可能となる。

本研究では上述した特徴を有するシステムを開発し、オスマン・トルコ語文書群を対象としたケーススタディを通じて、システムの有用性を検証する。

### 2. 開発したシステム

開発したシステムの概要を図 1 に示す。本システムは大きく、「Omeka を用いた画像とメタデータ管理」「TEI を用いた史料本文の記述」「可視化による史料の整理・分析支援」の 3 点から構成される。以下、それぞれについて説明する。

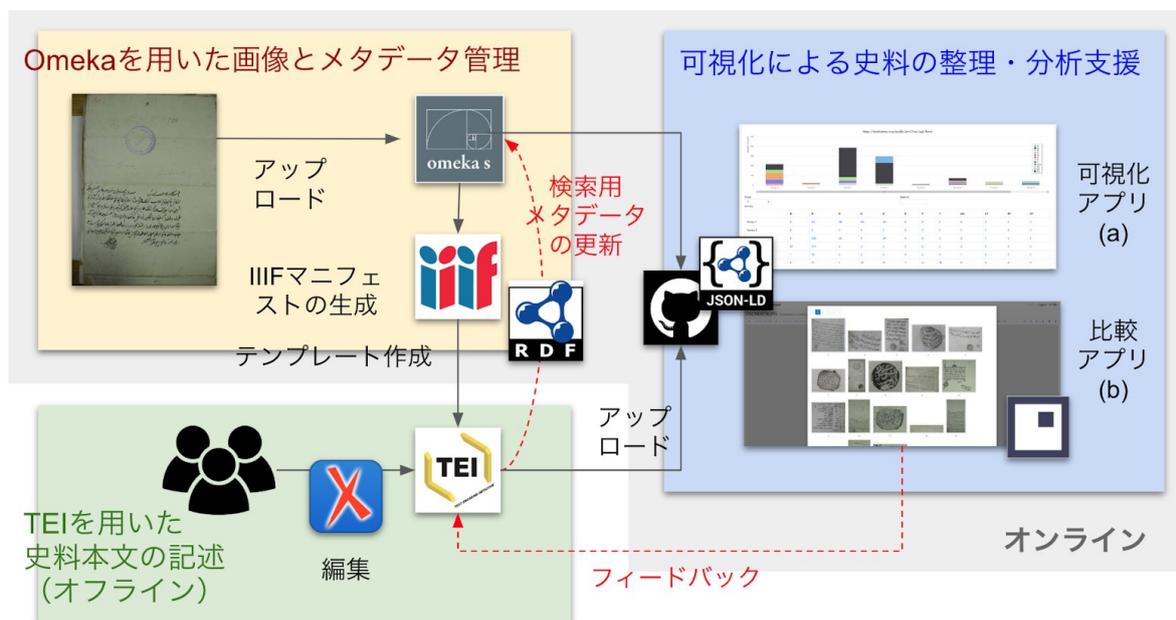


図 1 開発したシステム  
Figure 1 Overview of developed system.

## 2.1. Omeka を用いた画像とメタデータ管理 (オンライン)

まず Omeka S (以下, Omeka) に画像を登録し, 画像データを IIF 準拠で利用可能とする[7]. また対象史料に関する情報 (メタデータ) を登録し, それらに基づく史料の検索や管理などを行う. Omeka に登録したデータは非公開設定にすることで, 限定ユーザ間 (ログインユーザ) のみのデータ共有を行う.

次に Omeka の API を用いて IIF マニフェスト URI やキャンバス URI, 画像 URI などを抽出する. これらの情報を利用して, facsimile 要素にこれらの値を持つ TEI ファイルを生成する. 具体的なマークアップ方法については, TEI-C 東アジア/日本語分科会が作成しているガイドライン[8]を参考にした.

## 2.2. TEI を用いた史料本文の記述 (オフライン)

先に作成した TEI ファイルおよび Oxygen を利用して, 史料本文の記述を行う. 使用する要素については, 史料の特性および研究目的に応じて定める. また, Omeka から機械的に作成した facsimile 要素, および Oxygen の編集機能を用いることで, 画像中の特定領域とテキストデータを紐付けて管理可能な環境を提供する.

編集した TEI ファイルは GitHub の Private リポジトリを用いて管理することにより, ファイルのバージョン管理に加え, 限定ユーザ (コラボレーター) 間におけるデータ共有を行う. さらに,

Omeka が提供するデータ更新 API を利用することで, TEI で記述した情報を Omeka のメタデータとして登録する. これにより, TEI で記述した情報に基づく史料の検索を可能とする.

## 2.3. 可視化による史料の整理・分析支援 (オンライン)

Omeka のデータ取得 API を用いて抽出したメタデータを利用して, 特定の条件に基づく史料数の可視化 (図 1a) 等を行う[1]. さらに, TEI を導入することによって追加された機能として, ICP (IIF Curation Platform) [9]をベースとした部分画像比較機能を提供する「IIF Curation Comparison」

(図 1b)を開発した. TEI の facs 属性で指定された領域を使って自動的に作成した IIF キュレーションリストを入力として, メタデータに基づくキュレーションの絞り込みや, 検索結果の比較など可能とする. 例えば, 図 2 は CODH が提供する「大職冠」のキュレーション[10]を表示した例を示す. ファセットを用いて「身分」を「僧侶」に絞り込み, 選択したキュレーションを Mirador で比較している. なお, メタデータに基づくキュレーションの絞り込み等は IIF Curation Finder でも実現可能である. 本ツールの差分としては, Mirador を用いたキュレーション画像の比較ができる, デモ版についてはセットアップなしにブラウザ上で IIF Curation List を読み込み動作させることができる, 等が挙げられる.

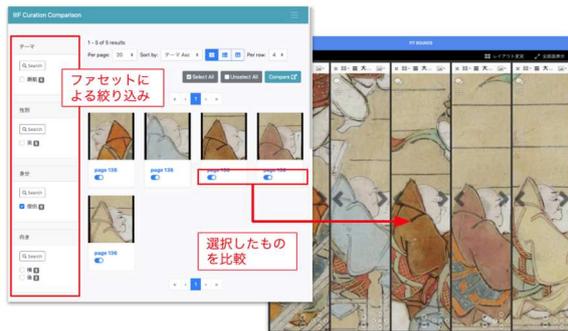


図 2 IIF Curation Comparison の画面例  
Figure 2 User interface of IIF Curation Comparison.

以上のように、TEI で記述したテキストデータと、Omeka に格納したメタデータを関連づけて可視化することにより、研究者への気づきの提案や、仮説の検証支援を行う。

### 3. ケーススタディ

本研究では、ケーススタディとしてオスマン・トルコ語文書の画像データ 2,268 点の整理および分析に対して、本システムを適用する[7]。これまで Omeka を用いて画像データを管理し、これらの画像データ群を 1,089 点の文書に分類している<sup>1</sup>。さらに、各文書のレイアウトや当該文書群が元々格納されていたフォルダに関する情報等をメタデータとして付与し、対象文書群の整理を進めている。以下では、ケーススタディの背景と目的を述べた後、TEI を用いた史料本文の記述と、可視化による史料の整理・分析の実例を示す。

#### 3.1. ケーススタディの背景と目的

本ケーススタディが対象とする史料は、ボスニア・ヘルツェゴヴィナ所在のクレシェヴォ修道院に所蔵されている、オスマン・トルコ語文書を中心とした 1,089 点の史料群である。

カトリックのフランシスコ修道会に属するクレシェヴォ修道院は、1463 年から 1878 年まで約 400 年間、オスマン帝国の支配下にあった。その間、管轄区信徒の日常的な行政手続きや修道院運営に関して、修道院とオスマン当局との間で文書が取り交わされ保管されたが、本史料群はその一部である。1765 年に修道院が火災に見舞われ多くの所蔵史料が焼失したため、残存文書は 18-19 世紀のものを中心とすると見られる。

本史料群の特徴として、これまで本格的な整理・読解が完了しておらず、管見の限り新出史料であることが挙げられる。史料が未整理のまま伝来した背景として、1878 年にクレシェヴォがオ

スマン帝国の支配を離れて以降、同修道院におけるオスマン・トルコ語利用が衰退し、本史料群を扱える修道士が不在となったことが考えられる。しかしながら本史料群は、従来研究が手薄になってきたオスマン帝国下のカトリック教徒の生活やフランシスコ会の役割を明らかにする可能性を秘めている。さらに、紛失・散逸を免れ地方に伝来しているという点でも貴重であり、帝都イスタンブール保管文書を主史料としてきたオスマン史研究に新たな知見を提供する可能性を有している。このため筆者らは、クレシェヴォ修道院から本史料群の画像データ 2,268 点を入手し、体系的な整理・分析に取り掛かった。

1,089 点の文書は、現在 9 つのフォルダに分けられ、所蔵されている。そのうち、本研究においてはフォルダ 1,2 に含まれる文書 198 点 (Tursko1 179 点, Tursko2 19 点) を検討対象とする。

具体的な研究アプローチとして、①画像データの管理とメタデータの作成による史料整理、②史料本文の記述と情報抽出による史料読解、③それらに基づく史料分析という 3 つの作業を行う。これらの作業において、2 章で述べたシステムを利用する。史料整理についてはこれまでの取り組みの中ですでに述べたため[7]、以下では、3.2 において史料読解について、3.3 において史料分析について述べる。

#### 3.2. 史料読解

史料読解では、読解の基礎となる史料本文の記述と、史料の整理・分析に必要な情報の抽出に取り組む。ここでの課題は、史料本文の記述において、本文のみならず、オスマン・トルコ語文書のもつ独自の様式をも記述することである。

本史料群の多くは、修道院とオスマン当局との間で取り交わされた行政文書であり、文書の機能ごとに一定の記述様式を持つ。ここで言う様式とは、文書の構成要素 (上奏・上申書・命令書・証書など)、各構成要素の配置・書体・大きさ・方向・インクの色などを指す。1 枚の文書が、複数の部局や差出人の間で回覧され、それぞれの主体が記述を行うこともあり、その場合、様式は図 3 のようにさらに複雑化する。このように様式は、文書の機能や作成プロセスを反映しているため、様式に関する情報を本文記述作業の中で保持することを課題とした。

上記課題を踏まえ、本研究では TEI を用いて、まず文書様式を記述し、その様式の枠組みに沿って史料本文を記述し、記述した史料本文にマークアップを施し情報抽出するという手順をとる。

<sup>1</sup> 本稿で示す文書点数は、3.3 の分析結果も含め、すべて Omeka を用いた分類に基づいている。しかしながら、これは暫定的な数字である。筆者らは、本史料群について 2,268 点の画像データを入手したが、現在ま

で文書そのものに触れて点数を確認する機会に恵まれていないためである。今後クレシェヴォ修道院で史料調査を行い、実際の文書点数を確認する予定である。

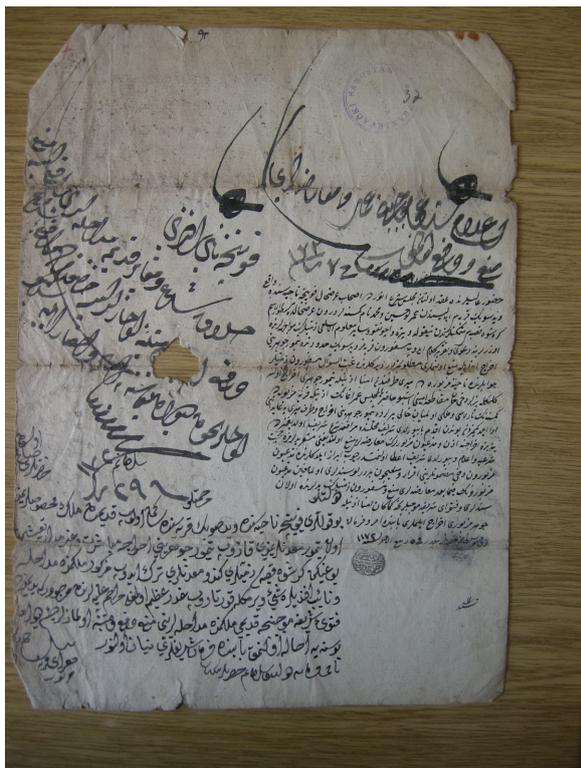


図 3 複数の構成要素をもつ文書  
Figure 3 Document with multiple components.

```

TEI facsimile
37 <text>
38 <body>
39 <div type="i'lām" facs="#zone_r11_3tv_cjb">
40 <p>
41 <lb/>... </lb/>
42 </p>
43 <div type="sign" facs="#zone_ojj_ktv_cjb">
44 <p>
45 <lb/>... </lb/>
46 </p>
47 </div>
48 </div>
49 <div type="buyuruldu" facs="#zone_gsh_ltv_cjb">
50 <p>
51 <lb/>... </lb/>
52 </p>
53 <div type="sign" facs="#zone_rb2_mtv_cjb">
54 <p>
55 <lb/>... </lb/>
56 </p>
57 <div type="stamp" facs="#zone_drk_ntv_cjb">
58 <p>
59 <lb/>... </lb/>
60 </p>
61 </div>
62 </div>
63 </div>
64 </body>
65 </text>
    
```

図 4 三階層構造の枠組み  
Figure 4 3 levels division of components.

### 3.2.1. TEI/XML ファイルの作成

TEI マークアップにあたり、2.2 で述べた通り、まず Omeka に登録した IIF 画像データに基づき、本文情報を記述するためのテンプレートとなる TEI/XML ファイルを作成する。teiHeader 要素には、Omeka に登録したアイテムのタイトルを与えた。また、facsimile 要素には、当該アイテムの IIF マニフェストに基づく情報を与える。body 要素は空の p タグを与える。これらを Omeka の API を用いて機械的に行い、計 1,089 件の XML ファイルを作成し、GitHub にアップロードした。

### 3.2.2. 様式の記述

空の body 要素内には、文書の構成要素を単位として、div タグによる三階層構造の枠組みを記述する。さらに、facs 属性を用いて記述箇所と画像データ中の特定領域とを紐づける。

例えば、図 4 と図 5 に示す事例は、大きく 2 つの構成要素から成る。裁判官が作成した上申書 (i'lām, 画像領域①) と、上申書を受けて州総督が作成した命令書 (buyuruldu, 画像領域③) である。これを第一階層とする。第一階層の構成要素は、さらに下位の構成要素として、それぞれ作成者を示す署名を持つ (sign, 画像領域②, ④)。これを第二階層とする。命令についてはさらに、署名の横に作成者の印影がある (stamp, 画像領域⑤)。これを第三階層とする。

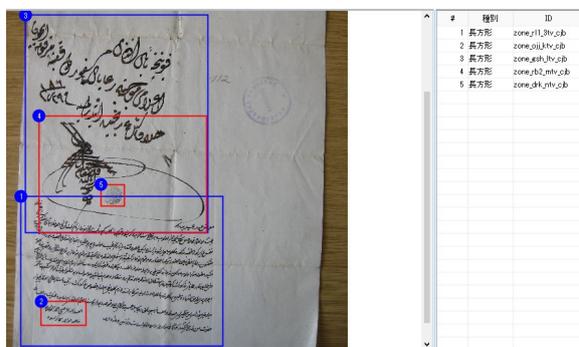


図 5 各構成要素の領域指定  
Figure 5 Zone definition for each component.

このような構成要素の三階層構造を、div タグの階層化によって表現する。構成要素の定義には、type 属性を用いる。これにより、各構成要素同士の関係性と文書全体における各部分の位置付けを明確にする。さらに、facs 属性を用いて画像データと対応させることで、各構成要素の配置・書体・大きさ・方向・インクの色・筆跡などの視覚情報を保持する。以上の作業により、文書様式に関する情報を記述した上で、史料本文の記述に移る。

### 3.2.3. 史料本文の記述

史料本文は、階層化された文書様式の枠組みと、文書自体の改行に従って、`lb` タグを用いて記述する。オスマン・トルコ語は本来アラビア文字で記述されるが、読み手の解釈を示すことを目的にラテン文字に転写することが慣例となっているため、本研究においてもラテン文字で本文記述を行う。

### 3.2.4. 情報の抽出：固有表現の記述

記述した本文に TEI マークアップを行い、史料の整理・分析に必要な情報を抽出する。現段階では基礎情報の収集を目的とし、作成日と作成者についてマークアップ作業を行う。なお、判読が困難な箇所は、`cert` 属性を用いて判読の確度を記録し、今後の作業の中で随時修正を加えられるようにした。

#### 3.2.4.1. 作成日のマークアップ

本文書群に記述された日付の性質は、単一ではなく、大きく分けて3種類ある。オスマン・トルコ語で記述された文書作成日の日付、オスマン・トルコ語で記述された作成日以外の日付（事件の発生日、過去の事柄への言及など）、クレシェヴォ修道院が文書を受領・保管した際に内容を要約してクロアチア語で記述した日付である。さらに暦も、ヒジュラ暦と西暦が入り混じっている。そのため、これらの情報を峻別し作成日として分析対象とするものを指定することが必要である。一方で、その他の情報も本文情報として保持する必要がある。

そこで本研究では、一次的な情報であると考えられるオスマン・トルコ語で記述されたヒジュラ暦の作成日を分析対象として指定することとし、峻別のために以下の3つの記述ルール(表 1)を定め、`date` タグと `when` 属性を用いてマークアップを行う。

第一に、一次的な情報であるオスマン・トルコ語と二次的な情報であるクロアチア語を区別するための記述ルールとして、`type` 属性を用い、クロアチア語要約に `type="summary_Croatian"` を付す。第二に、ヒジュラ暦と西暦を区別するための記述ルールとして、ヒジュラ暦には `when-custom` 属性を、西暦には `when` 属性を用いる。第三に、作成日とそれ以外の日付を区別するための記述ルールとして、`type` 属性を用い、作成日に `type="created"` を付す。

以上3つの記述ルールに従い、分析対象を指定して情報抽出するとともに、その他の情報の保持に努めた。なお、文書中に作成日が記述されていないものは `unknown` とし、作成月日が不明なものは暫定的に作成年末の12月30日を入力する。以上の作業により抽出された作成日は、分析の段階で `Hutime` の暦変換サービス[11]を用いて西暦に変換する。

表 1 日付のマークアップ記述

Table 1 Markup rules for date.

日付の性質	マークアップ記述例
ota/H/作成日 (分析対象)	<code>&lt;date when-custom="1186-04-29" type="created"&gt;</code>
ota/H/作成日以外	<code>&lt;date when-custom="1186-04-29"&gt;</code>
hr/H	<code>&lt;date when-custom="1186-04-29" type="summary_Croatian"&gt;</code>
hr/M	<code>&lt;date when="1772-07-30" type="summary_Croatian"&gt;</code>

【凡例】 ota：オスマン・トルコ語， hr：クロアチア語， H：ヒジュラ暦， M：西暦

#### 3.2.4.2. 作成者のマークアップ

文書の作成者は、署名と印影によって示されている。しかしながら、必ずしも常に署名と印影の両方が揃っているわけではなく、また、署名の筆跡の乱れや印影のかすれ・欠けによって、判読困難な場合もある。そのため、作成者の特定には、署名と印影の画像サンプルを蓄積し、画像を比較しながら判読確度を高めることが重要となる。

そこで本研究では、作成者の特定とマークアップの準備作業として、まず署名と印影の画像サンプルの抽出・蓄積を課題とした。具体的な作業として、これらの構成要素を `div` タグの第2階層・第3階層として枠組みを作成し、図 5 に示したように、`facs` 属性を用いて画像領域と紐づけた。判読・特定まで至ったものについては、`persName` タグと `role` 属性を用い、名前と役職をマークアップする。現段階で、マークアップによる情報抽出まで進めたサンプルは少なく、今後鋭意作業を行う予定である。

今後の展望として、署名や印影箇所の自動抽出や、切り出した画像の類似度に基づく判読支援等、本文の作成およびマークアップ作業の半自動化または自動化のための手段も検討する。

### 3.3. 可視化機能を使った分析例

ここでは、3.2 で整理した情報、および2. で述べた可視化機能を用いた分析例について述べる。以下、3.2.2 で述べた「様式に基づく分析」と、3.2.4 で述べた「固有表現を用いた分析」に分けて説明する。

#### 3.3.1. 様式に基づく分析

3.2.2 でマークアップした階層構造のうち、第一階層毎の文書数を可視化した。その結果を図 6 に示す。また文書数の詳細な内訳を、図 7 の Total 列に示す。なお、1枚の文書中に複数の構成要素が含まれる場合があるため、ここでの数字は、3.1 で示した文書点数より多くなっている。

図 6 から、文書の構成要素として、「`hüccet`」「`buyuruldu`」「`i'lām`」の順に多いことが視覚的に確認できる。「`hüccet`」は証書、「`buyuruldu`」は

命令書, 「i'lām」は上申書を指す。これらは, 従来オスマン帝国史研究において非常によく利用されてきた「シャリーア法定台帳」と呼ばれる史料の作成過程を示す, 重要な手がかりである。

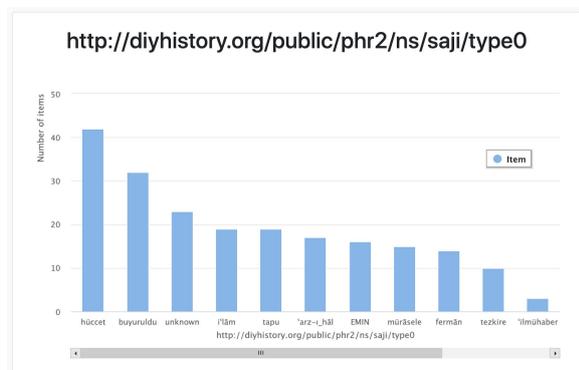


図 6 第一階層毎の文書数

Figure 6 Number of documents by the first component.

オスマン帝国各地には, カーディー (裁判官) あるいはその代理であるナーイブが主催するシャリーア法廷があり, 司法のみならず地方行政の中心でもあった。「シャリーア法廷台帳」とは, カーディー (ナーイブ) が作成あるいは受領した文書の控えを集めた帳簿である<sup>2</sup>。冊子形態を取り多数伝世しているが, 一方, 帳簿作成のために参照された文書は, 紙片であり, カーディーから各宛先に向けて送付されるものは, 散逸する性質をもつ。

本史料群が比較的多く含んでいる構成要素は, まさにこの散逸傾向にある文書である。これらの検討は, シャリーア法廷台帳の作成過程やシャリーア法廷の機能の見直しにつながる可能性を有していると言える。また, 本来法廷に保管されるべき文書が現在修道院に保管されている点も, 修道院の機能を考える上で興味深い。

さらに, 他の要素 (メタデータ) である「フォルダ」と第一階層の関係性を可視化した。具体的には, 開発したシステムの可視化機能を利用し, 図 6 で示した可視化結果に対して, さらに「フォルダ」毎の文書数を可視化した。その結果を図 7, 縦軸と横軸を反転させたものを図 8 に示す。

本史料群は, 3.1 で述べたように, 現在 9 つのフォルダに分けられ保管されているが, これらのフォルダが何に基づいて分類・形成されたのか現段階では不明であり, 今後の検討課題の一つとなっている。そのため, フォルダ毎に構成要素やその他の項目の分布を分析することにより, 修道院がフォルダを形成した背景や, 修道院にとっての個々の文書の意味付けを探る狙いがある。

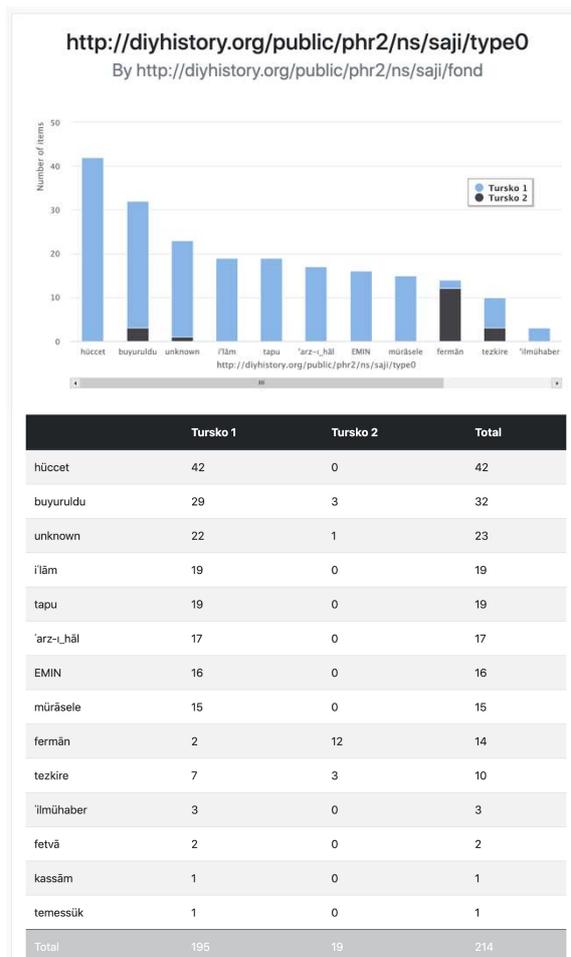


図 7 第一階層毎のフォルダ別文書数

Figure 7 Number of documents by the first component and document folder.

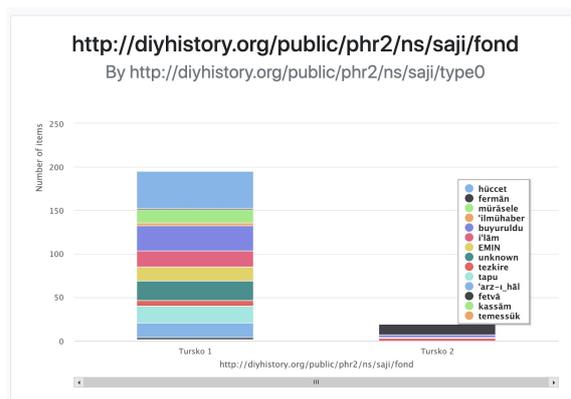


図 8 フォルダ毎の第一階層別文書数

Figure 8 Number of documents by the first component and document folder.

図 7 と図 8 から, フォルダ 1 とフォルダ 2 は, 質的に大きく異なることが視覚的に確認できる。

<sup>2</sup> シャリーア法定台帳 (Sicil) . <http://tbias.jp/ottomansources/sicil>

まず、含まれている文書点数が、フォルダ1は多くフォルダ2は少ない。構成要素の種類についても同様である。すなわち、フォルダ1は比較的雑多な種類の文書が多数収められているのに対し、フォルダ2は特定種類の文書が少数収められていると言える。さらに、「fermān」の収蔵状況は特徴的である。「fermān」とは、オスマン帝国君主によって下された勅令のことであり<sup>3</sup>、重要度の高い文書である。図からは「fermān」について、所蔵点数は少数であるものの、ほとんどがフォルダ2に収められていることが分かる。このことは、フォルダ2が、修道院にとって重要な少数の文書を集めて形成されたフォルダである可能性を示唆している。この分析に基づき、今後の作業ではフォルダ2の検討を優先的に行う方針を固めた。また、このようなフォルダ間の性質の差について今後の作業の中で十分注意する必要があることを認識した。

### 3.3.2. 固有表現を用いた分析

3.2.4でマークアップした固有表現のうち、まず作成年月日を用いた文書数の時系列変化を可視化した。その結果を図9に示す。なお、可視化にあたって、作成年月日を年に集約している。

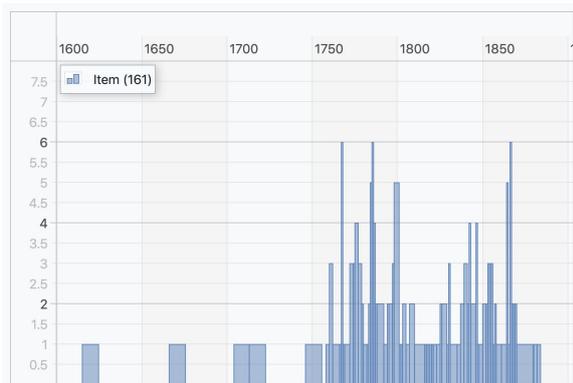


図9 文書数の時系列変化

Figure 9 Time series change of the number of documents.

作成年は、3.1で述べたように、1765年の修道院火災から1878年のオスマン帝国支配終結までを中心とすると見られた。

図9から、中心となるのはやはり1760-1880年の文書であることが検証された。この期間中の分布はばらついているが、今後さらに情報を集積し、

特徴的な年の分析を行う予定である。一方で、少数ながら、火災以前あるいはオスマン支配終結後の文書も含まれていることが分かった。現段階で最も古い文書は1615年、最も新しい文書は1882年に作成されたものである。これらの文書からは、情報の少ない火災以前、あるいはハプスブルク支配に入った後のクレシェヴォ修道院の様子が明らかにできる可能性がある。

次に、作成者に基づく分析を行う。具体的には、図2で述べたIIIF Curation Comparisonを利用して、TEIファイルから作成したIIIF Curation Listを読み込み、マークアップした印影を比較する。その比較結果の画面例を図10に示す。

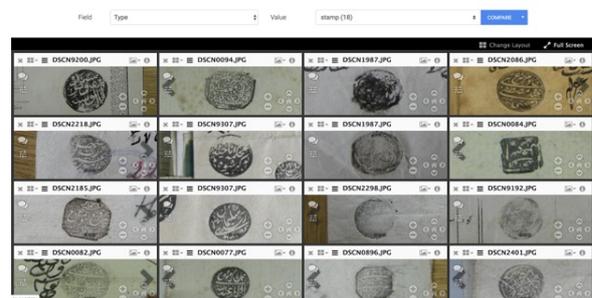


図10 ICPを用いた印影の比較例

Figure 10 Comparison of stamps with ICP.

図10からは、ここに挙げた印影が全て異なっていること、つまり文書の作成者が全員異なる人物であることが分かる。このことは、クレシェヴォ修道院の運営や管轄区信徒の日常生活が、多くのオスマン役人の支える文書行政の中で営まれていたのだという気付きを与えてくれた。作成者の特定には至らなかったが、今後画像サンプルをさらに蓄積し、同一人物に属する署名と印影の特定を進める予定である。

またその後の展望として、同一人物によって作成された文書のリスト化と、同一年に属する文書の作成者のリスト化を行う。前者リストは、各作成者の任期・職掌・クレシェヴォ修道院との関係性の分析、および特定人物に着目した地方文書行政の分析に有用である。後者リストは、一定期間にどのような組織によってクレシェヴォ修道院に関する文書行政が担われたかを分析するために有用である。

<sup>3</sup> 本稿で使用した「fermān」という用語は、本来単独では勅令を意味せず、「命令」と訳すことが適切である。オスマン・トルコ語の原語では、勅令を指す用語として「emr-i şerif」または「emr-i 'ālī」という表現が用いられる(高松洋一、オスマン朝のハットウ・ヒューマーンについての一考察：切り取られたハットウ・ヒューマーンの検討を中心に、東洋文化, No.91,

pp.101-145, 2011.)。しかしながら、先行研究においては勅令を指して「fermān」の呼称を用いることが一般化しており、またクレシェヴォ修道院で作成されたクロアチア語要約の分類名も「fermān」とされていることから、現段階ではこの用語を構成要素名として使用した。今後適切な用語を検討する。

## 4. 考察

### 4.1. 歴史研究的観点による考察

本システムを利用した史料整理と分析について、3点の有用性が確認された。

第一に、同一システム上で史料の整理・読解・分析という3つの作業を行うことにより、個々の作業の進展と、作業同士の連携との両立が可能となった。従来段階的に進めていた作業プロセスが同時並行的に進められるようになったことは、作業効率の向上につながった。たとえば、可視化により早い段階で史料の性格や傾向を把握し、それに基づいて読解作業の方針を決定することは、効率化に有用であった。多量の史料を扱う場合、史料読解の段階で膨大な時間を要し、史料に没入するあまり全体像の把握が遅れ、効率的な研究計画が立てられないことがあるが、本システムは史料読解と分析が同時に行えるため、常に個々の文書と史料群の全体像を行き来しながら作業を進めることができた。第二に、TEIを用いることにより、オスマン・トルコ語文書のもつ様式情報を保持することが可能となった。従来、様式情報は重要であっても記述が困難なため史料本文に反映させることが難しかったが、div タグの構造化により、複雑なものであっても再現することができた。簡便な記述方法でありながら、画像データと紐づけることにより視覚情報を失わず保持できる点も優れている。第三に、高い検索機能により、容易に特定の入力情報に基づいて文書を取り出し一覧化することが可能となった。具体的には、メタデータ・様式・固有表現・史料本文・それらの情報の確度・作業記録など、様々なレベルで必要な情報を取り出すことができるため、特定の構成要素をもつ文書や特定年に作成された文書の集中的な分析ができた。また確度の低い情報のみを取り出して見直し作業を行ったり、作業記録に基づいて再撮影が必要な文書をリスト化したりすることができた。

### 4.2. システム的観点による考察

本研究では、オンラインとオフラインを併合した研究環境を提案した。具体的には、TEIを用いた史料本文の記述環境として、Oxygenを利用した。既存のTEI編集用オンラインツールとの差分として、TEI文書の検証、イメージマップエディターを用いた画像中の特定領域とテキストデータの紐付けを容易に実現することができた。特に、後者の機能を用いた作成したデータを利用することにより、IIIF Curation APIを用いた印影の比較等が可能となった。

また、オンライン環境として、IIIF準拠の画像データを公開するためのOmeka、主にTEI文書を共有するためのGitHubを利用した。Omekaを用いて画像をIIIF準拠で利用可能とすることにより、先述したICPを用いたデータ分析環境を容易

に(IIIF準拠の既存ツールをベースとして)実現することが可能となった。特に、TEIとIIIF Curation APIの接続可能性を示した点については、DH研究の発展に貢献し得る本研究の成果の一つであると考えられる。

## 5. 結論

本研究では、TEIとIIIFを用いて歴史学研究者の史料整理と分析を支援するためのシステムを開発した。さらに既存の関連ツール群を最大限利用すること、および研究者の作業環境を考慮し、オンラインとオフラインを併合した研究環境を提案した。TEIを用いた史料情報の記述、および可視化による史料の整理・分析支援を行うことで、歴史学研究者に対して新たな気づきを提示し得ることを確認した。

今後はケーススタディで述べた史料研究を進めるとともに、本研究で開発したオンライン環境(Omeka, GitHub, 可視化ツール)とオフライン環境(Oxygen)の接続部分のAPI化など、汎用的に利用可能な環境の構築を目指す。

## 参考文献

- [1] 中村覚, 大和裕幸, 稗方和夫, 満行泰河, 鈴木淳, 吉田ますみ. Linked Data を用いた歴史学研究者の史料管理と活用を支援するシステムの開発, 情報処理学会論文誌, Vol.59, No.2, pp.267-277, 2018.
- [2] みんなで翻刻. <https://honkoku.org/>, (参照 2019-10-28)
- [3] FromThePage. <https://fromthepage.com/>, (参照 2019-10-28)
- [4] Recogito. <https://recogito.pelagios.org/>, (参照 2019-10-28)
- [5] Omeka. <https://omeka.org/>, (参照 2019-10-28)
- [6] Oxygen. <https://www.oxygenxml.com/>, (参照 2019-10-28)
- [7] 佐治奈通子, 中村覚. 歴史学と情報学のより良い協働を目指して—オープンなDH支援ツールを用いたボスニアのカトリック修道院所蔵のオスマン・トルコ語文書群のデータ整理の一事例, 研究報告人文科学とコンピュータ (CH), Vol.2019-CH-120, No.11, pp.1-7, 2019.
- [8] TEI-C 東アジア/日本語分科会. IIIF 画像とのリンク,