

歴史学と情報学のより良い協働を目指してー 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

佐治奈通子^{†1} 中村覚^{†1}

概要：本発表では、歴史学と情報学の協働による、史料画像データ整理の実践事例を示す。具体的には、歴史史料から得られる情報を整理・分析可能な史料研究支援システムを利用して、ボスニアのカトリック修道院所蔵の未整理のオスマン・トルコ語文書の画像データ 2,268 点を整理する。その作業を通じて、歴史学的な観点からのニーズを反映させたシステムの改良を試みた。また、個々のデータ詳述と、可視化による史料群の全体像の把握が可能となったことで、作業の効率化とバランスのよい史料理解に繋がった。

キーワード：オスマン朝、ボスニア、フランシスコ会、修道院文書、史料整理、Linked Data、Omeka S

Toward better collaboration between history and informatics - An example of arrangement of Ottoman-Turkish documents from the Catholic monastery in Bosnia with open DH support tools

NATSUKO SAJI^{†1} SATORU NAKAMURA^{†1}

1. はじめに

本発表では、歴史学と情報学の協働による、史料画像データ整理の実践事例を示す。具体的には、歴史史料から得られる情報を整理・分析可能な史料研究支援システム[1]を利用して、ボスニアのカトリック修道院所蔵の未整理のオスマン・トルコ語文書の画像データ 2,268 点を整理する。その作業を通じて、歴史学的な観点からのニーズを反映させたシステムの改良を試みる。これらを踏まえ、文書整理・史料研究から得られた知見、それらのプロセスにおけるシステムの有用性について考察する。

2. 史料について

ここでは、本発表で扱うボスニアのカトリック修道院所蔵のオスマン・トルコ語文書群について説明する。

2.1 史料の概要

本文書群は、14 世紀末に創設されたカトリックのフランシスコ会修道院であるクレシェヴォ修道院に保管されている、オスマン・トルコ語文書を中心とした史料群である。クレシェヴォ修道院は、ボスニア・ヘルツェゴヴィナ中央

ボスニア県クレシェヴォに位置しており、同地が 15 世紀後半にオスマン朝支配を受けて以来、多数のオスマン・トルコ語文書が作成・保管されてきた。クレシェヴォ修道院付属の図書館には、本文書群の他にも、医学・薬学関係の書物を中心に、様々な言語で書かれた約 22,000 点の刊本・手稿本・文書類などが収蔵されており[2]、未整理のものも含めると約 50,000 点の所蔵資料があるという。しかしながら、現在オスマン・トルコ語を解する修道士がいないため、所蔵資料のうちオスマン・トルコ語のものの多くは未整理のままになっている。本文書群もそのひとつである。

文書群は、9 つのフォルダ[a]から成っている。クレシェヴォ修道院が火災に見舞われ多くの所蔵資料が焼失したとみられる 1765 年から、クレシェヴォがオスマン朝支配を離れた 19 世紀後半までに作成されたオスマン・トルコ語文書が中心であると推測される。未整理ながら、フォルダ分け・一部整理番号の割当て・一部梗概の作成（手書き・タイピングなど複数種類）・写真撮影などがすでに行われており、これまで複数回に亘って史料整理が試みられてきた形跡が見られる。

^{†1} 東京大学
The University of Tokyo.

保存されていた。この単位に対し特定の呼び名は史料入手段階では付されていなかったため、本研究ではこの単位を「フォルダ」とする。

a) 文書群は 9 つの単位に分けられ、図 1 に示すような紙製のフォルダに

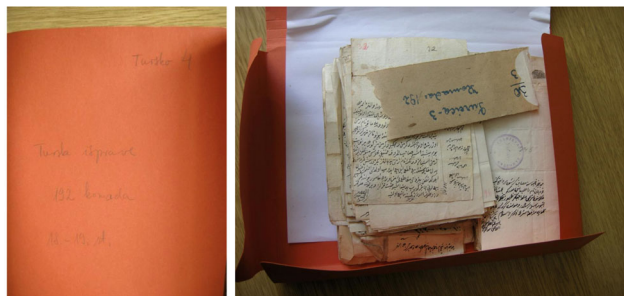


図 1 フォルダの例
Figure 1 Example of folder.

発表者は 2015 年 2 月、撮影が完了していたこれら文書の画像データ 2,268 点を CD-R で受領し、文書整理を依頼された。文書整理にあたり、史料群の性格記述などアーカイブ学的な観点と、記述内容・年代分析など歴史学的な観点の双方から、総合的な史料整理が必要であると考えた。

2.2 史料の重要性

ここでは、本文書群の歴史学的な価値について検討し、整理作業の意義や展望を提示する。発表者は、1. オスマン朝下のカトリック教徒の状況を示す史料であること、2. 地方にまとまって伝世していた史料であること、3. 新出史料であることの 3 点から、本文書群の整理には歴史学的意義があると考えた。

2.2.1 オスマン朝下のカトリック教徒

オスマン朝領域では、様々な宗派共同体に属する人々が共存しており、非ムスリムについては従来、「ミット制」のもとでの一体性を持った宗派共同体像が描かれてきた。実はこれが歴史的に形成されたものであったことが近年の研究で明らかにされてきたが[b]、この古典的な「ミット制」理解のためか、オスマン支配下の非ムスリムについては、ギリシア正教徒・アルメニア教徒・ユダヤ教徒の宗派共同体の存在が大前提とされ、カトリック教徒の研究は看過されがちであった。本文書群は、カトリック修道院に伝世していた文書群であることから、これまで研究史上手薄であったオスマン朝下のカトリック教徒の状況の一端を明らかにする可能性のある史料であると考えた。

2.2.2 オスマン朝の文書行政

オスマン朝では、非常に文書行政が発達していた。文書は、イスタンブールの中央政府だけでなく帝国各地で作成され、やり取りされた。文書の内容・種類は多岐に亘っており、また複雑な文書行政の中では、文書とともにそれらの

内容を写し取った控え台帳が作成されたり、君主をはじめとする複数の責任者が文書を回覧して、承認の印を付したり必要事項を新たに書き込んだりした[c]。

このようにして大量に作成された文書は、トルコ共和国イスタンブールの大統領府オスマン文書館（T.C. Cumhurbaşkanlığı Devlet Arşivleri Başkanlığı Osmanlı Arşivi, 以下 BOA）に収集・保管されており、BOA の所蔵資料を利用した研究はオスマン史の中心のひとつとなっている。しかしながら、BOA の所蔵史料の中心となっているのは、中央のイスタンブールに残存していた文書である。一方、地方に残存していた文書は、現地に保存されているか、紛失・焼失・売却など様々な理由ですでに失われている場合が多く、特にボスニア・ヘルツェゴヴィナについては、先のボスニア紛争で焼失してしまったものも多い[3]。そうした状況の中、本文書群は、現地にまとまった形で残存していたという点で非常に貴重な史料群である。また、BOA の所蔵史料は複数回に亘る史料整理の中で出所原則が乱されてしまったものもあり[4]、それに比して出所が明確であるという点でも重要である。本文書群は、地方に伝世した文書の側から、オスマン朝の文書行政を再検討するために有用であると考えた。

2.2.3 BOA での史料調査

発表者は、2019 年 1-3 月、BOA で史料調査を行った。クレシェヴォやクレシェヴォ修道院に関する史料を中心に調査・収集を行ったが、その中で本文書群に含まれる文書と同一の史料の所蔵は確認できなかった。また、管見の限り本文書群を利用した研究も見当たらなかった。無論さらなる調査・検討が必要であるが、現時点では、本文書群が新出史料であり、この整理・分析から得られる情報はこれまでに知られていなかったものである可能性が高いと考える。

以上 3 点より、本文書群の整理作業には十分な歴史的意義があるものと考えた。今後、アーカイブ学的・歴史学的観点の双方から整理を試み、得られた情報を分析・提示していきたい。

3. 史料研究支援システムについて

ここでは本研究で使用する史料研究支援システムについて説明する。後述する章において、歴史研究者のニーズに基づくシステムの改良を行うため、ここでは本システムの改修前における基本的な機能について説明する。

b) 「ミット制」についての議論や研究動向については、上野雅由樹、ミット制研究とオスマン帝国下の非ムスリム共同体、2010、史学雑誌、119 (11)、pp. 1870-1887 に詳しい。

c) オスマン朝の文書行政については、高松洋一、オスマン朝における文書・帳簿の作成と保存：18 世紀から 19 世紀初頭を中心に、史資料ハブ地域文化研究、2004、no. 4、pp. 106-126 に詳しく説明されている。

本システムの概要を図 2 に示す。本システムはウェブアプリケーションであり、史料研究プロセスを構成する史料の「収集」「整理」「分析」を支援する機能群を提供する。

システムの実装については、Java を用いて開発してきた従来のシステムから、Omeka S（以下、Omeka）をベースとしたシステムに移行した[5]。

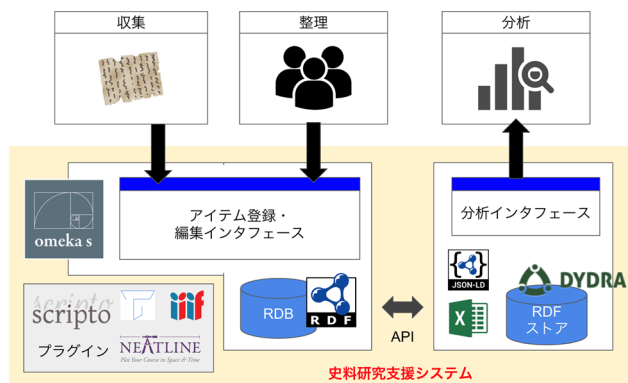


図 2 開発した史料研究支援システムの概要

Figure 2 System overview.

3.1 収集プロセス

「収集」プロセスにおいては、Omeka が提供する登録インタフェースを用いて、画像等のメディアを登録する。登録されたメディアはモジュール「IIIF Server」の機能により、IIIF が定める API（IIIF Image API・IIIF Presentation API）に準拠して利用することが可能となる。また登録されたメタデータは RDF 形式で格納され、それぞれのアイテムやメディアに URI が与えられる。

3.2 整理プロセス

「整理」プロセスにおいては、Omeka が提供する編集インタフェースを用いて、登録されたアイテムやメディアのメタデータを編集することにより、史料の書誌的な情報や、研究メモ等を登録する。メタデータの入力に関しては、モジュール「RDF Datatype」による日付や整数などの値の型指定、モジュール「Custom Vocab」による用語リストの作成による表記揺れの吸収、などの機能追加を行なっている。メタデータのプロパティについては、Dublin Core 等の既存の語彙を登録可能なほか、モジュール「Custom Ontology」による研究者の目的に応じたプロパティ（メタデータ項目）定義を可能としている。さらにメタデータの編集機能に加え、モジュール「Scripto」による翻刻が可能となっている。

3.3 分析プロセス

「分析」プロセスにおいては Omeka の API を利用し、システムに格納された RDF データを Microsoft Excel や JSON-

LD 等の各種フォーマットでダウンロードし、史料の並び替えや比較等の分析を行う。特に、JSON-LD 形式のデータに対しては、プロパティ毎の史料数の可視化図や、史料数の時系列変化の可視化図を閲覧可能な機能を提供する。これらの可視化図を利用することにより、史料の整理と並行して、史料の大局的な傾向の把握を支援する。さらに詳細な分析が必要な場合には、出力した RDF データを Dydra 等の RDF ストアに登録し、SPARQL Endpoint を利用した史料抽出、それに基づく史料分析を行う。

4. 作業プロセス

実施した作業プロセスは以下の通りである。本研究では、文書そのものではなく、すでにデジタル化された画像データを基に文書整理を試みたことを、作業の前提として強調しておきたい。

1. 画像データ入手
2. 画像ファイル名を ID として、画像単位の整理
3. 整理項目の検討
 - a. エクセルへ全般的な情報入力
 - b. 共通事項・分析事項のあぶり出し
 - c. 暫定的な整理項目の決定
4. システムへのスキーマ登録、画像の一括登録
5. システムへの入力・可視化
 - a. アスペクト比を基にした文書分類
 - b. 可視化による直感的な全体像の把握
6. 入力作業からのフィードバック・システムの改良

大まかな作業プロセスは以上の通りである。以下では、「6. 入力作業からのフィードバック・システムの改良」について詳述する。

また、本プロセスを通じて、本稿執筆時に採用している整理項目（スキーマ）の例を表 1 に示す。以下のスキーマをアイテムの登録・編集時のテンプレートとして利用することにより、データ入力時の作業効率化や表記ゆれの吸収を行う。なお、接頭名詞空間 ex は、本研究において独自に定義したものである。

表 1 整理項目（スキーマ）の例

Table 1 Example of properties.

プロパティ名	URI	Range（値域）
タイトル	dc:terms:title	Literal（文字列）
サイズ	ex:aspectRatioType	カスタム語彙（A, B, C, ...）から選択
言語	ex:lang	カスタム語彙（bos, ota, ...）から選択
フォルダ	ex:folder	カスタム語彙（Tursko 1, Tursko 2, ...）から選択
オスマン朝スタンプの個数	ex:stamp_o	xsd:integer（整数）
メモ	ex:memo	Literal（文字列）

4.1 システムに対するニーズ

システムを利用した史料整理において、歴史研究者からのニーズは以下の通りである。

- 新出史料として、一から文書整理を行う
- アーカイブズ学的な情報と、歴史的な情報の双方を記述する
- 入力情報を可視化する
- 画像を意味のあるまとまりにわけ
- 画像同士の関係性を記述する
- フォルダごとの分析を行う
- スタンプ・定型句・人名・地名などの表記を比較する
- ヒジラ暦から西暦に自動変換する
- 文書同士の関係性を記述する
- 地名と緯度経度情報を結びつけてマッピングする (GIS)

これらのニーズを受け、本研究では、1. 画像同士の関係性の整理、2. フォルダ別の可視化、3. アノテーションの付与に取り組んだ。以下ではこの3点について、より詳しく要望の内容と背景を説明する。

4.1.1 画像同士の関係性の整理

前述の通り、本研究で対象とした文書群は本研究の開始前から写真撮影によるデジタル化が完了していた。しかしながら、この画像データ群は本来史料整理に不可欠であるデジタル撮影についてのドキュメンテーションがほとんど行われておらず、撮影ルール（撮影順番・撮影箇所など）や撮影時の文書の保管状態・原型・スケールなどが不明であるという問題があった。画像データは全部で 2,268 点だが、1 枚の文書について、表裏・全体像・一部分のアップ画像など複数枚が撮影されており、それらの画像データの連なりからは、一見して元の 1 枚の文書の姿がイメージしづらく、文書の総数も不明であった。

以上のことより、まずは実物の文書を扱うように、画像単位ではなく文書単位での整理が必要である。また、画像同士の関係性を記述することで、画像が文書のどの部分にあたるのかを把握する必要がある。

4.1.2 フォルダ別の可視化

前述の通り、この文書群にはすでに複数回整理が試みられた形跡があった。梗概が作成されたことからみても、過去にはオスマン語を解する人物が携わり、何らかの根拠をもって史料整理を行ったことが推察される。しかしながら、現在編まれている 9 つのフォルダについては、どのように分類・成立したのかが不明である。

これらのフォルダの成立過程を考えることは、アーカイブズ学的に重要であるのみならず、歴史的にも、クレシェヴォ修道院にとっての各文書の意味付けや重要性を考える手がかりになる。そのため、フォルダ別に入力データを可視化し、過去の文書整理作業の痕跡と併せて分析することで、各フォルダに含まれている文書の特徴を分析することが求められる。

4.1.3 アノテーションの付与

オスマン朝の文書は、多くが行政文書であることから一定の記述様式を持っている。その様式を理解することによって、文書の種類を特定したり、どのような手順で作成されたのかという文書行政の在り方を検討したりすることが可能となる。ここで言う様式には、文書の構成要素（花押・陳述文・定型句・承認印など）、各構成要素の配置、文字の書体・方向・色などの情報を含む。これらの情報を詳しく記述することによって、ボトムアップ式に同じ様式を持つ文書を分類したり、構成要素毎の比較・分析をしたりすることが可能となると考えた。

4.2 対応

4.1 の歴史研究者のニーズに基づき、システムに加えた改良点について述べる。

4.2.1 画像同士の関係性の整理

画像同士の関係性の整理、具体的には、複数の画像から構成されるまとまり（これを「アイテム」と呼ぶ）の管理は、Omeka の標準機能で対応可能である。したがって、本研究では、この標準機能を用いて画像をアイテム単位に整理した上で、システムに登録する方法を採用した。

一方、このような画像群をアイテム単位に整理していく作業に対しては、Tropy[d]の利用が考えられる。Tropy とは、文書館等で撮影した史料画像やデジタルアーカイブで公開

d) “Tropy” . <https://trophy.org/>. (参照 2019-04-08).

されているデジタル化史料等をローカルで管理するためのソフトウェアである。

Tropy はローカルでの作業内容を Omeka S にエクスポートする機能を提供するため、3 で述べたシステムに登録する前作業として、Tropy での画像整理を行うことを検討した。しかし、Omeka S へのエクスポート機能は 2019 年 4 月時点においてベータ版であり[e]、例えばアイテムの新規登録はできるが更新ができない、等の一部機能の不足が確認された。この結果、本研究では Tropy の導入を見送った。クラウド版 Tropy の開発[f]を含め、今後のシステム改良に期待したい。

4.2.2 フォルダ別の可視化

本ニーズに対しては、3.3 で述べたプロパティ毎の史料の出現頻度グラフに対して、さらに特定のプロパティの値毎に分類（色分け）する可視化機能を追加した。

図 2 に例を示す。上図が前者の可視化図であり、プロパティ「フォルダ (ex:folder)」の値 (Tursko 1, Tursko 2, ...) 毎の史料数を表示している。下図が改良を加えた可視化図であり、前者の各棒グラフの値について、さらにフィールド「サイズ (ex:aspectRatioType)」の値に基づき分類（色付け）を行なっている。この分類項目（プロパティ）はユーザが任意に選択することができる。

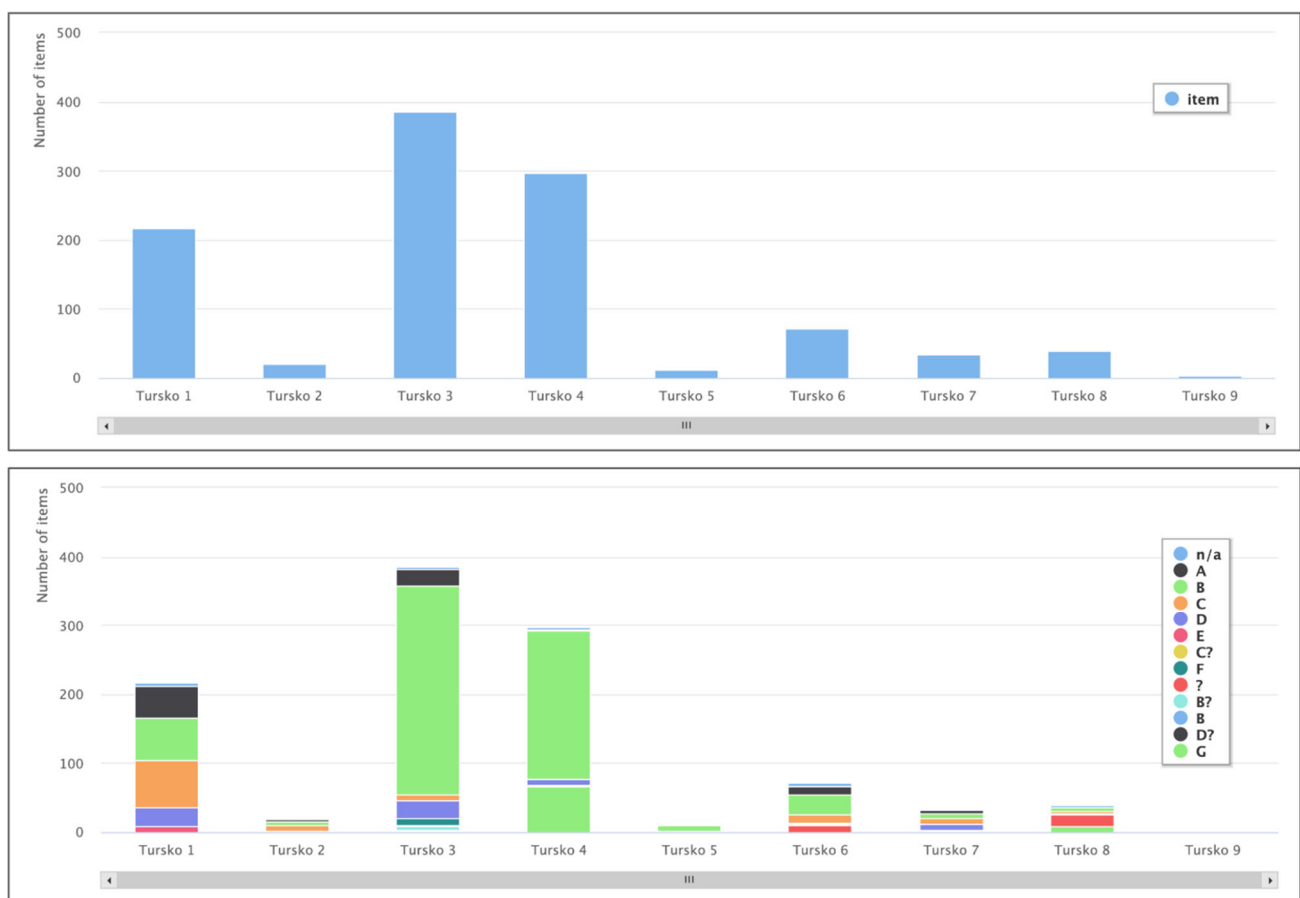


図 3 フォルダ別のアイテム数の可視化（サイズ別の色付けを含む）

Figure 3 Visualization of the number of items by folder and size.

4.2.3 アノテーションの付与

4.1.3 のニーズに対して、文書整理に必要な情報の粒度と Omeka の標準機能の対応を図 4 に示す。メディア（画像等）、複数のメディアから構成されるアイテム、複数のアイ

テムから構成されるアイテムセット（コレクション）の管理については、Omeka が提供する標準機能を利用することができる。

一方、史料中のスタンプに関する情報の記述、等の画像

e) “Export to Omeka S - Documentation”, <https://docs.tropy.org/other-features/omeka/>, (参照 2019-04-08).

f) “Next Steps for Tropy”, <https://tropy.org/blog/next-steps/>, (参照 2019-04-

中の特定の箇所に関する記述を行うアノテーションについては、Omekaの標準機能としては提供されていない。

そのため、モジュール「Annotate」等を利用し、画像に対するアノテーション付与機能を追加した。本モジュールはWeb Annotation Ontologyに基づき、Omekaに登録済みのメディアに対するアノテーション付与を可能とする。

本モジュールを用いることで、図4に示す4つの粒度のデータを3で述べたシステム上で管理可能とした。なお、これらはすべてLinked Dataとして関連づけられるため、これらの関係を用いたデータ分析（スタンプのみの抽出・比較等）が可能である。

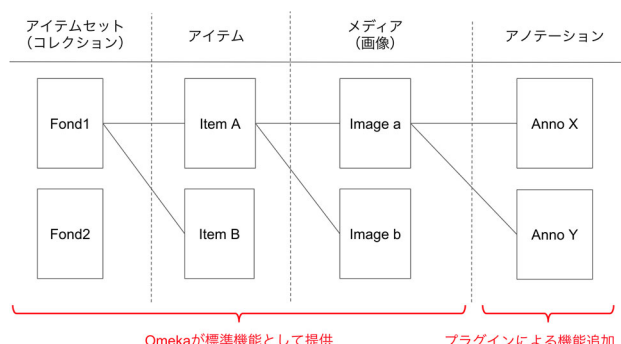


図4 文書整理に必要な情報の粒度とOmekaの機能
Figure 4 Hierarchy required to document arrangement and functions of Omeka.

4.3 結果

以上のシステム改良を受け、得られた結果について述べる。

4.3.1 画像同士の関係性の整理

システムの標準機能の利用によるものではあるが、画像をアイテムに整理することにより、画像単位ではなく文書単位の整理が可能になった。それにより、当初分からなかった実際の文書点数を確認することができた。画像2,268点に対し、文書は1,089点である。現状、フォルダ表紙など文書以外の画像を1アイテムとして登録しており、また1画像中に6枚の文書が含まれている事例もあるため、1,089点という数字は暫定的なものである。今後の作業の中でさらに正確な点数確認を行う予定である。アイテム内の画像同士の関係性の記述は、メディア（画像）単位のメタデータを付与することで可能になった。また、複数の画像を1アイテムに整理した上でデータ入力を行うことは、画像ひとつひとつについて作業するよりも手数が減るため、効率が上がり、入力ミスが少なくなるという効果も得られた。

4.3.2 フォルダ別の可視化

可視化によって、フォルダごとの文書の収蔵状況の直感的な把握が容易となった。また、選択した分類項目によってさらに詳しく各フォルダ内の状況を分析できるようになった。本研究では、特にアスペクト比に基づくサイズ分類（A・B・C・D・E・F・G）を行い、それぞれのサイズの文書がどのフォルダ内に収蔵されているのかを可視化した。フォルダ全体を見ると、Tursko 1, Tursko 3, Tursko 4に文書の大半が収められている。逆に、非常に点数の少ないその他のフォルダには、何らか特定の機能や特徴を持った文書が収められている可能性が考えられる。また各フォルダの内訳について見ると、Tursko 3, Tursko 4の多くはサイズBの文書から構成されている。これを受け今後、サイズBの文書を中心に、年代・構成要素・人名・地名などの情報の入力作業を進めていく予定である。

4.3.3 アノテーションの付与

アノテーション機能を利用し、文書様式について構成要素毎の詳細な記述を行った。図5に例を示す。画像に直接情報を付与できるため、内容のみならず、構成要素の持つ位置・形・大きさなどの情報の記述も可能となった。今後、同様の方法でアノテーション付与作業を継続してサンプルを集積し、構成要素を用いた文書の様式分類や、構成要素毎の比較を行う。

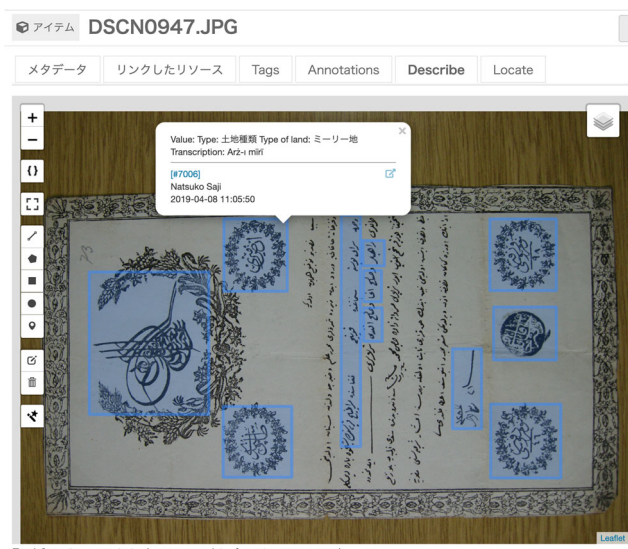


図5 アノテーションの付与例

Figure 5 Example of annotations.

なお、本稿執筆時点においては、システム上での視認性の観点から、アノテーションの値を「要素A: 値a」「要素B: 値b」という形式で記述している。再利用を意図した記述ルールを定めているが、非構造化データとして管理して

いる。一方、システム上では各アノテーションに URI が与えられ、各アノテーションのメタデータを管理することが可能である。今後の取り組みとして、アノテーションを記述するためのスキーマを定め、記述内容の構造化を行う。これにより、アノテーション単位での比較分析を容易とする。さらにアイテムのメタデータにも反映させることにより、アノテーションの情報に基づく史料検索を可能とする。

5. まとめ

本発表では、歴史学と情報学の協働による、史料画像データ整理の実践事例を示した。具体的には、歴史史料から得られる情報を整理・分析可能な史料研究支援システムを利用して、ボスニアのカトリック修道院所蔵の未整理のオスマン・トルコ語文書の画像データ 2,268 点の整理に取り組んだ。本事例を通じて、以下に示すような歴史学、情報学、それぞれにおける成果が得られた。

歴史学的な成果としては、画像データ・アーカイブズ学的な情報・歴史学的な情報を、すべてひとつのシステム上に入力して管理することが可能になった点が挙げられる。これまでは画像管理・史料群の性格記述・史料読解を別々のソフトウェアで行ってきたため、それぞれの情報をリンクさせ総合的に分析する機会自体が失われていた。しかし、本システムの利用により情報が統合され、検索・可視化機能によって総合的に分析することが可能となった。また、情報入力と同時に可視化図が作成されることは、史料の示す情報を客観的に把握することに繋がった。通常、史料整理・読解作業はインプットに長い時間を要し全体像を見失いがちであるが、本システムを利用することで、各文書の詳細と文書群の全体像を容易に往復することが可能となった。

情報学的な成果としては、歴史研究者からのフィードバックに基づき、システムの各種改良を行うことができた点が挙げられる。可視化図の改良やアノテーション機能の追加は、研究対象や領域を問わず利用可能な機能群であり、他の研究事例に本システムを利用した際の効果が期待できる。

本研究の実施期間においては、歴史学研究者からのニーズの提示、情報学研究者からのソリューションの提示、というやりとりを週 1 回程度の頻度で実施した。歴史学と情報学の協働にあたっては、綿密なコミュニケーションが特に重要であった。双方の評価環境の違いなどの困難さは存在するが、このような学際的な取り組みが、それぞれの専門分野における新しい成果の創出に寄与し得ることを確認した。

謝辞

本研究を行うにあたって、東京大学大学院人文社会系研究科人文情報学拠点人文情報学概論の授業において、下田正弘先生をはじめとする先生方にご指導いただいた。ここに感謝の意を表する。また、クレシェヴォ修道院の文書整理に携わる機会を与えて下さった、清水美穂さん、江川ひかり先生に感謝の意を表する。

参考文献

- [1] 中村寛, 大和裕幸, 稗方和夫, 満行泰河, 鈴木淳, 吉田ますみ. Linked Data を用いた歴史研究者の史料管理と活用を支援するシステムの開発, 2018, vol.59, no.2, pp.267-277.
- [2] Skrbo, Armin, Izet Masic, Selma Skrbo, Elma Ramakic, and Lejla Zunic. Bosnian Franciscans and the Monasteries in Kresevo and Fojnica as Source of Scientific Bibliography, *Materia Socio-Medica*, 2017, vol.29, no.2, pp.149-154.
- [3] 佐治奈通子. “サラエボ東洋学研究所.” In 世界の図書館から: アジア研究のための図書館・公文書館ガイド = From the libraries of the world: a guide to libraries and archives for Asian studies, edited by U-PARL, 勉誠出版, 2019, pp.186-187.
- [4] 高松洋一. オスマン朝における文書・帳簿の作成と保存: 18 世紀から 19 世紀初頭を中心に, 史資料ハブ地域文化研究, 2004, no.4, pp.106-126.
- [5] 中村寛. Omeka S を用いた研究データ管理システムの試作. 人文情報学月報, 2018, vol.89.