

Linked Data 化した典拠データと IIF を用いた法帖の異版比較支援システムの開発

中村 覚（東京大学 情報基盤センター）

成田 健太郎（埼玉大学 大学院人文社会科学研究科）

永井 正勝（東京大学 附属図書館）

法帖の研究において、版の異同を検討し、その系統を詳らかにすることは肝要である。本研究ではこのような法帖研究の支援を目的として、法帖における異版作品の比較を支援するシステムを開発した。典拠データの Linked Data 化と、複数機関が提供する IIF 準拠画像に対するアノテーション付与による個別作品の識別により、異版関係にある個別作品を検出可能なシステムを構築した。国立国会図書館、国文学研究資料館、東京大学附属図書館が提供する法帖画像を対象としたケーススタディを通じ、版 6 件、著者 252 件、作品 902 件、異版作品 1403 件から構成される典拠データの作成、および 213 件の個別作品に対して紐づけを行い、任意の異版作品の異なる個別資料における再現例を検出できることを確認した。

Development of System to Support Comparison of Different Editions of Calligraphic Models with Linked Data of Authorities and IIF

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

Kentaro Narita (Graduate School of Humanities and Social Sciences, Saitama University)

Masakatsu Nagai (University of Tokyo Library System)

In studies of calligraphic models, it is vital to review differences in edition and to recognize their lines. In this research, for the purpose of supporting such research, we developed a system to support comparison of different editions of calligraphic models. This system supports to detect individual works in different editions by Linked Data of authority and identify individual works by giving annotations to IIF compliant images provided by multiple institutions. Through a case study which uses the images of calligraphic models from several institutions, we confirmed that it can detect reproduction examples of a certain different edition work in different items.

1. はじめに

法帖（ほうじょう）とは、本来主に紙に書かれていた書跡を石材表面に転写して刻し、それから採拓した複製物を指す。それらは書道の手本ないし美術品として伝えられ、書道史や文化史の研究資料となっている。法帖には、一つの作品から成る単帖（たんじょう）、一人の筆者の複数の作品を取めた専帖（せんじょう）、複数の筆者の作品を取めた集帖（しゅうじょう）の区別がある。

同一の法帖の刊刻は一回のみとは限らず、原石の損傷や滅失、あるいは需要増大等の理由により、拓本に基づいて翻刻されることは少なくない（ここで言う翻刻とは、手書きの文字を活字に起こすという意味ではなく、ある版をその原初になるべく近い形で複製または復元するという意味である）。その場合、一つの法帖に異なる原石から採拓された異本が存在することになり、それらと比較検討することは法帖研究の重要なターゲットとなる。

また、複数の法帖が同一の作品を収録していることも多い。そのような場合、収録されている法帖ごとに同一の作品が異なる属性を有し、その比較検討もまた研究上重要な課題となる。本研究では便宜上、ある法帖は編集行為によって版（edition）として確定するものと考え、その後の翻刻によって生じる異本については版の異なりと認めない。というのも、以下参照する先行研究では、編集行為の異なりを単位として法帖を弁別しており、当面これを利用することで本研究の実現可能性が確保できるからである。将来的には、本研究を発展させる形で、翻刻によって生じる異本も記述できるシステムの実現が望まれる。

法帖に関する先行研究に、版ごとに作品を記述し整理したものはあるが、それによって各作品と異版関係にある他の作品を参照することはできない。法帖のデジタル化・オープン化が進展したときには、同版で異なる個別資料に体现された作品（以下、「個別作品」）を一覧できることだけ

でなく、異版関係にある個別作品も一覧できることが望まれる。

したがって本研究では、まず、版ごとに作品を記述した先行研究を利用して、当座の典拠データを作成する。さらに、それを個別作品と紐づけることにより、異版関係にある個別作品を検出し、それらの比較を支援するシステムの開発を目的とする。

2. 法帖研究の手順と課題

法帖研究において、たとえば、複数の版に収録されている可能性のある作品を詳細に検討する場合、一般に以下のような手順を要する。

- ①当該作品を収録している版を先行研究によって確認する。
- ②当該版の拓本資料の所在を各種目録（紙媒体・ウェブ）によって確認する。
- ③所蔵機関に赴いて個別資料を閲覧し、必要に応じて撮影する。
- ④収集した情報を整理し、当該作品の異なる版における再現例を比較する。

①の手順では、たとえば先行研究[1]によって、各版が収録している作品を知ることは容易となっているものの、反対に各作品がどの版に収録されているかを漏れなく把握することは容易でない。②の手順にも改善すべき課題があるが、本研究では対象としない。③の手順は、個人の努力による効率化に限界があり、研究者にとって大きな負担となる。④の手順において、個々の研究者は撮影した画像データを整理して保存し、個々の作品がどの版のどの部分に収録されているかといった知見を、たとえば画像データを加工する、紙にプリントアウトして書き込みを施す、別途表形式データファイルを作成するなど、何らかの形で記述していると思われるが、それらが研究者のコミュニティにおいて共有されることは通常ない。人文学におけるオープンデータ化で望まれるのは、研究者個人の経験や判断が大きく影響する④の手順であると考えられる。

法帖に対する包括的研究は、従来大量の法帖を収集した個々の研究者がそれらを手元に置いて繰り返し検討し、得られた知見を十分に整理して紙媒体で発表する流れを常態としてきたが、そのような研究が可能なほど資料へのアクセシビリティに恵まれている研究者は少ないのが現実である。しかし今後、デジタルアーカイブの発展によって、資料へのアクセシビリティは向上する可能性があり、またデジタル技術は研究者間で専門的知見を共有し、膨大な作業を分担するツールとして有用性を増していくと思われる。すなわち、個々の研究者が研究の過程で得た知見が典拠データ上に集積され、法帖資料群に対する包括的理解が徐々に豊かになり、常時最新の状態で参照可

能となっているという研究モデルが構想されるのである。

3. システム構築手順

異版比較支援システムの構築手順を図1に示す。「A. 典拠データの Linked Data 化」「B. アノテーションによる個別作品の識別」「C. 法帖の検索・比較アプリケーション開発」の3つのプロセスから構成される。

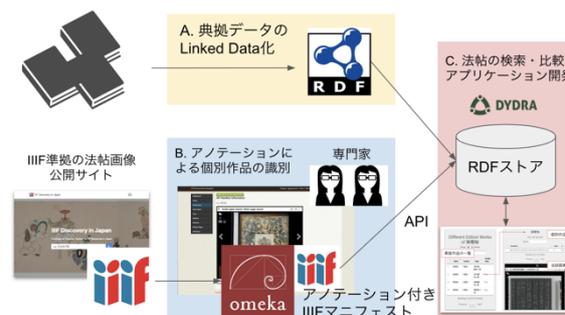


図1 システム構築手順
Figure 1 Flow of system development.

3.1. 典拠データの Linked Data 化

3.1.1. 異版作品について

一般に法帖には多くの作品が収録され、同一の作品が異なる法帖に収録されていることも多い。先行研究[1]は、各種法帖に収録されている作品名をすべて書き出した著作であり、その法帖-作品の組（以下、「異版作品」）に対して識別子（以下、「ID」）を付与し、当座の典拠データを得る。

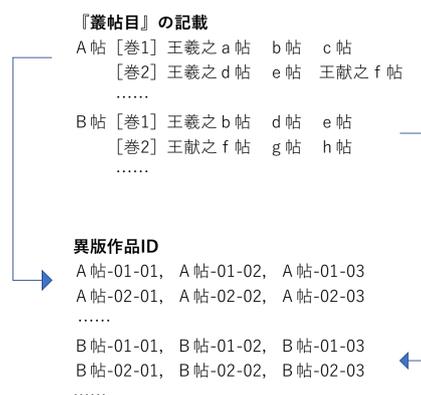


図2 異版作品 ID
Figure 2 Different edition work ID.

図2ではb帖、d帖、f帖がA帖、B帖のいずれにも見え、一見すると両者に共通のIDを付与すべきであるように思えるが、異なる作品に偶然同じ名称が与えられている同名異帖の可能性もあり、作品としての同一性を検証するためには当該法帖を閲覧するなどの作業を要する。このような作業は、歴史的資料を扱う人文学の基礎的研究

行為ではあるが、2の手順④に記したように、デジタル画像が入手できたとしても、データ整理そのものが個々の研究者に委ねられるため、判断の根拠が可視化・共有されることが通常はない。本研究の目的の1つは、原資料から作業仮説を立ち上げるプロセスを研究者コミュニティに開き、思考とデータの共有化を図ることにある。もちろん将来的には作品ごとにIDが付与されるのが望ましいが、本研究では当面の作業の便宜のため、異版作品ごとにIDを付与する。

3.1.2 典拠データの Linked Data 化

この異版作品IDに基づき、典拠データの Linked Data 化を行う。構築する典拠データの例を図3上部に示す。

クラスとして、「版」「作品」「人物」「異版作品」クラスを用意する。先行研究[1]を当座の典拠データとして捉え、そこに含まれる各作品や版をインスタンスとして記述する。インスタンスのURIについては、「版」「作品」「人物」クラスのインスタンスには語彙「dbpedia-ja」, 「異版作品」クラスのインスタンスには独自語彙「ex」を用いた。

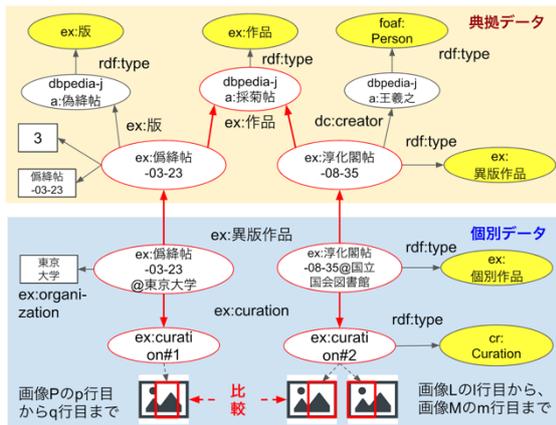


図3 典拠データと個別データのRDF表現
Figure 3 RDF representation of authority data and individual data.

3.2. アノテーションによる個別作品の識別

異版作品IDを異なる個別資料の画像データに結びつければ、異版作品IDを介して、ある異版作品の異なる個別資料における再現例(以下「個別作品」という)を一挙に検出することが可能になる。

異版作品IDと画像データを結びつけるにあたっては、画像の位置情報にIDを付与し、このIDを異版作品IDに結びつける。これにより、ある作品の範囲がどの行からどの行までであるかという専門家の知見を共有し、ある作品について範囲を特定して個別作品画像を提供することが可能になる。

より具体的には、以下の「アノテーション付与」「アノテーションと典拠データの紐付け」の2つのプロセスを踏む。

3.2.1 アノテーションの付与

本プロセスでは、ウェブ上で画像が公開されている個別資料中から、個別作品を識別することを目的とする。

個別作品の識別については、中村ら[2]が提案したシステムを利用する。具体的には、Omeka Classicをベースシステムとして、プラグインの一つである「IIIF Toolkit with Mirador[3]」を用いる。本システムはウェブ上で公開されている個別資料のIIIFマニフェストをシステムに読み込むことで、個別資料の各画像をシステムにインポートする機能を提供する。さらに、読み込んだ各画像に対して、画像中の範囲を選択して、アノテーションを付与する機能を提供する。本研究では、書道研究者が本システムを実際に利用することで、個別資料の各画像中に含まれる各個別作品の範囲をアノテーションによって識別した。本インタフェースを図4に示す。

具体的には、画像中の個別作品の範囲を矩形で選択し、異版作品IDをテキストデータとして入力した。なお、1つの個別作品が複数のページにまたがる場合には、各個別作品に対して複数のアノテーションが紐付けられる。

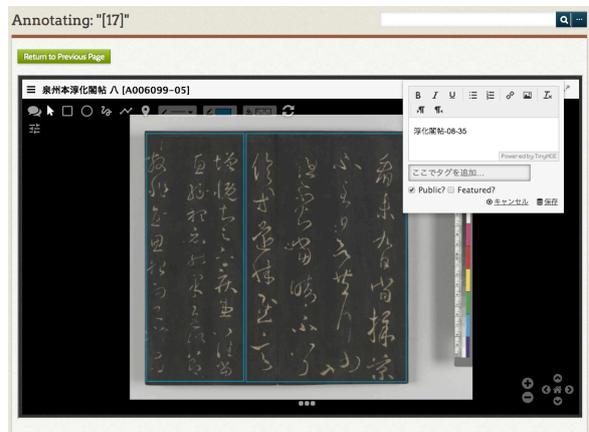


図4 アノテーション付与インタフェース
Figure 4 Interface to edit annotations.

3.2.2. アノテーションと典拠データの関連付け

この結果、個別作品毎の範囲が指定されたアノテーション情報を含むIIIFマニフェストが生成される。これを利用することにより、oa:FragmentSelectorとしてアノテーションの位置情報を取得できるほか、対象となるCanvas IDやManifest URI, 異版作品ID, 提供機関名(Attribution)が得られる。この異版作品IDと提供機関名を加えたIDに基づくURIを用いて、個別作品毎にインスタンスを作成する。

アノテーションに関する情報 (Manifest URI, Canvas ID, 位置情報) の個別作品インスタンスへの紐付けにおいては, oa:Annotation クラスのインスタンスを cr:Curation クラスのインスタンスに変換した, 紐付けを行なった. この理由として, IIF Curation API に準拠して記述することにより, アノテーションとして選択した複数の部分矩形領域について, 一意の URI を与え, さらに順序付きリストとして保持することが可能となる. この URI を個別作品に紐づけることにより, 個別データの管理を簡略化できることに加え, IIF Curation Viewer 等の各種ソフトウェアを利用可能になるといった副次的な利点が存在する.

3.3. 法帖の検索・比較アプリケーション開発

3.1 で作成した典拠データと 3.2 で作成した個別データが関連づけられることにより, 典拠データに基づく個別作品の検索が可能となる. 例えば, 図 3 中において赤色で示したリンクを辿ることにより, 例えば作品「採菊帖」から版「偽絳帖」および「淳化閣帖」に含まれる異版作品を抽出し, それらが含まれる個別作品の該当箇所を識別するキュレーションのリストを取得することで, 当該画像を閲覧, 比較することが可能となる.

本プロセスでは, このような検索過程を実現するウェブアプリケーションを構築する. アプリケーション構築にあたっては, クラウド上の RDF トリプルストアのサービスである Dydra[4] を利用し, RDF データを格納する. そして, Dydra が提供する SPARQL Endpoint と ajax 通信によって接続する JavaScript アプリケーションを構築する. 構築したアプリケーションの画面例などについては, 次章にて説明する.

4. ケーススタディ

3 で述べた手法に基づき, 法帖の異版比較を支援するアプリケーション[5]を構築した.

典拠データについては, 本稿執筆時点において版 6 件, 著者 252 件, 作品 902 件, 異版作品 1403 件から構成されている. また, この異版作品に対して, 213 件の個別作品を紐づけ, これらの画像を閲覧することができる. これらは, 東京大学附属図書館, 国立国会図書館, 国文学研究資料館から収集した個別資料が含まれる. 提供機関を横断した IIF コンテンツの収集については, 中村らが開発した検索システム[6]を利用した. 例えば検索語「淳化閣帖」に対して, 国立国会図書館所蔵資料 12 件, 東京大学附属図書館所蔵資料 6 件が検索結果として得られる.

4.1. 異版作品の検索と比較

開発したアプリケーションを利用することにより, 異版関係にある個別作品を検出し, それらを比較可能することができるかを確認する. 具体

的には, 図 3 で示した関係を辿ることで, 作品別・著者別・版別に異版作品を検索する.

例えば, まず作品「採菊帖」の異版作品の一覧を取得する. この結果, 図 5 に示すように, 4 つの異版作品が抽出される. 本インタフェースでは, 異版作品が「作品名」「著者」「版」「巻」および「画像の有無」の項目とともに表示される.



図 5 作品「採菊帖」の異版作品の一覧表示画面例
Figure 5 Interface to display different edition works of “Cai ju tie”.

この各異版作品に対して, 図 6 に示すように, 個別作品の一覧が得られる. 本インタフェースでは, 先に述べた項目に加え, 対象個別作品の「提供機関」名が表示される. ここでは, 「国文学研究資料館」「国立国会図書館」「東京大学」の 3 機関が公開する作品「採菊帖」・版「淳化閣帖」の個別作品の一覧が表示されている. 各作品の閲覧には IIF Curation Viewer[7]を用いる.



図 6 作品「採菊帖」・版「淳化閣帖」の個別作品の一覧および閲覧画面例

Figure 6 Interface to display the work “Cai ju tie” and the edition “Chunhua Ge tie”.

さらに, 作品名に与えられたリンクから, 図 7 に示すように, 当該作品の異版作品における個別作品の一覧を得ることができる. 本画面では, 比較対象とする行を選択し, 画面上部の比較ボタンをクリックすることにより, 関連づけられた IIF のアノテーションの情報に基づき, 対象とする個別作品が登場するキャンバスおよび箇所がフォーカスされた形で Mirador 上に表示される. この例

では、作品「採菊帖」について、国立国会図書館から1件[8]、国文学研究資料館（静岡県立中央図書館）から1件[9]、東京大学から3件[10][11][12]収集した個別作品を表示している例を示す。



図 7 作品「採菊帖」の異版作品における個別作品の一覧および比較画面例

Figure 7 Interface to compare individual works in different editions of “Cai ju tie”.

以上より、本システムを利用することにより、本研究で目的とした、任意の作品の異なる版における再現例を一挙に検出、閲覧することが可能となったことを確認した。

4.2. 画像処理による異版参照

開発したシステムの応用例として、画像処理による異版参照を試みた。

本研究では、当座の典拠データとして利用した先行研究[1]にしたがって各作品の名称を決定し、それを手がかりとして異版作品を検出した。しかしながら、法帖の作品名は一意に決定されるものではないため、典拠データにおいて、同一作品の名称が版によって異なっている可能性（同帖異名）がある。それらに対する名寄せを目的として、画像中の任意の箇所を範囲選択し、画像処理技術によって他資料中の一致する部分を検出する。将来的には、同一作品と確認された異版作品に同一の作品 ID を付与し、作品ごとの典拠コントロールを実現することが期待される。

この画像処理による異版参照を実用化するための初歩的実験として、先に挙げた作品「採菊帖」を例にとり、異版作品3件、個別作品5件の画像を対象として、相互に類似度を評価した。具体的には、ある画像から特徴点を検出し、それらの特徴点を他の画像中から探索して計数し、その数値が大きいくほど類似度が高いと評価する。特徴点抽出には SIFT を利用し、特徴点の比較には Brute-Force（総当り法）を用いた。

この結果を表 1 に示す。また、特徴点比較の可視化例を図 8 に示す。値の大小の違いはあるが、同一作品間における類似性を確認することができた。今後は、対象作品を拡大させ、誤判定の有無を含めた検証、アルゴリズムの改良や各種パラメータの調整を行う。

表 1 特徴点の比較結果

Table 1 Comparison result of key points.

	淳化閣帖 @ NDL	淳化閣帖 @ 国文研	淳化閣帖 @ 東大	偽絳帖 @ 東大	星鳳楼帖 @ 東大
淳化閣帖 @ NDL	-	286	28	35	42
淳化閣帖 @ 国文研	262 (A)	-	22 (B)	16	19
淳化閣帖 @ 東大	13	13	-	14	13
偽絳帖 @ 東大	20	24	7	-	319
星鳳楼帖 @ 東大	26	30	8	335	-

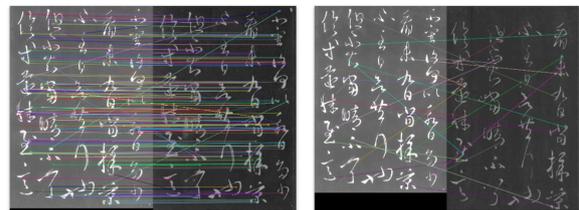


図 8 左: 表 1 (A), 右: 表 1 (B) における特徴点比較の可視化例

Figure 8 Visualization example of comparison of key points.

また、特に高い類似度を示した 2 対の個別作品について、考察を加える。

まず『淳化閣帖』について、国文学研究資料館（静岡県立中央図書館所蔵本を使用）および国立国会図書館は肃府本、東京大学附属図書館は泉州本と呼ばれ、異なる翻刻に係る異本である。前者 2 件は相互に高い類似度を示し、いずれも後者 1 件と低い類似度を示した。この結果は、本画像処理によって先行研究[1]が設定しているよりも細かい異本レベルの異同を定量的に把握できる可能性を示唆している。

さらに『星鳳楼帖』と『偽絳帖』は偽帖（より古い時代に制作された価値の高い法帖と詐称する、実際はより新しい法帖）であり、宇野雪村はいずれも明代の偽刻と推測しているが[13]、両版の偽刻の経緯について詳細な研究はなされていない。画像処理の評価を信頼するならば、この 2 件の類似度が排他的に高いという結果から、両版が系統関係の近い（たとえば共通の素材を用いて制作された）法帖である可能性が示唆される。

これらの結果、画像処理技術が法帖の近縁関係を判定するツールとして機能し得る可能性が示された。今後はより詳細な比較実験等を実施する予定である。

5. 考察

5.1. 開発したシステムの新規性・有用性

異版の比較支援を目的とした研究は、北本ら [14]による研究など、すでに数多くの取り組みが見られる。本研究の新規性の一つは、複数サイトから公開される IIF コンテンツを対象とすることでサイトを横断した比較を容易にできるようにしたという点である。2でも述べた通り、デジタルアーカイブ、IIFの発展によって、資料へのアクセシビリティは向上する可能性があり、またデジタル技術は研究者間で専門的知見を共有し、膨大な作業を分担するツールとして有用性を増していくと思われる。

5.2. アノテーションを用いた作品範囲の指定

本研究では、IIFのアノテーション機能を使用し、集帖中の作品を指定する方法を採用した。これは、各作品は基本的に行単位で区切られるためである。一方、例外的ではあるが、行途中で作品が分かれるケースも存在する。本研究では四角形でのアノテーションを前提としているが、今後は多角形による作品範囲の指定を可能にすることが求められる。

5.3. テキストデータ作成について

本研究では、4.1で述べた通り、画像にアノテーションを付与する方法を採用したが、TEI等に準拠したテキストデータの作成により、異版作品の絞り込みを実現する方法も考えられる。しかし法帖の場合、草書体などの読みにくい文字で記述されているため、専門家でも解読において意見が分かれることがある。そのため、テキストを確定する前に、異版を複数閲覧し比較検討する必要がある。文献としての利用を想定するとしても、その前段階として、資料からテキストを抽出する作業が必要であり、その作業の質を確保するためには複数の異版を手早く比較対照できる環境が望ましい。本研究は、この作業を支援するシステムとして位置付けている。

6. 結論

本研究では、法帖における異版作品の比較を支援するシステムを開発した。典拠データの Linked Data化と、IIF準拠画像に対するアノテーション付与による個別作品の識別により、異版関係にある個別作品を検出可能なシステムを構築した。

本システムが人文学に与える影響の一つは、個々の研究者が行っていた情報整理のプロセスおよび根拠となる資料を可視化させるとともに、アプリケーションでの検索というかたちで、整理の結果を共有させた点にある。これを可能としたのが IIF 準拠画像である。画像処理技術を利用した作品の同定支援システムの精度向上やテキストデータの TEI 化など取り組むべき課題も多いが、本研究は、IIF 準拠画像を積極的に活用した

人文知とデータの可視化の事例であり、歴史的資料の利用を伴う人文系学問におけるオープンデータ化（人文知のオープン化）の一つのモデルとして考える。

参考文献

- [1]. 容庚：叢帖目。中華書局香港分局（1980-1986）。
- [2]. 中村覚，成田健太郎，永井正勝，富澤かな：U-PARLにおける漢籍・碑帖拓本デジタルアーカイブの試作と研究利用。研究報告人文科学とコンピュータ（CH），Vol. 2018-CH-116, No. 5, pp. 1-8（2018）。
- [3]. Omeka Classic - IIF Toolkit, <https://omeka.org/classic/plugins/IiifItems/>, (参照 2018-10-31).
- [4]. Dydra, <https://dydra.com/>, (参照 2018-10-31).
- [5]. 異版比較支援プロトタイプシステム, <https://nakamura196.github.io/narita/>, (参照 2018-10-31).
- [6]. 中村覚，永崎研宣：日本国内の IIF 準拠画像に対する横断検索システムの構築，研究報告人文科学とコンピュータ（CH），Vol. 2018-CH-118, No. 8, pp. 1-6（2018）。
- [7]. 人文学オープンデータ共同利用センター，IIF Curation Viewer, <http://codh.rois.ac.jp/software/iif-curation-viewer/>, (参照 2018-10-31).
- [8]. 国立国会図書館デジタルコレクション，淳化閣帖第 1-10. [8], <http://dl.ndl.go.jp/info:ndljp/pid/2587034>, (参照 2018-10-31).
- [9]. 新日本古典籍総合データベース，淳化閣帖, <https://kotenseki.nijl.ac.jp/biblio/100065344>, (参照 2018-10-31).
- [10]. U-PARL 漢籍・碑帖拓本資料，泉州本淳化閣帖 八 [A006099-05], <https://iif.dl.itc.u-tokyo.ac.jp/repo/s/uparl/document/bff6f7ca-9187-4f44-aeda-89227316ab62>, (参照 2018-10-31).
- [11]. U-PARL 漢籍・碑帖拓本資料，偽絳帖 三 [A005936-03], <https://iif.dl.itc.u-tokyo.ac.jp/repo/s/uparl/document/f58ea819-fffb-4fcb-889c-61a332e0474a>, (参照 2018-10-31).
- [12]. U-PARL 漢籍・碑帖拓本資料，星鳳樓帖 卯 [A005935-04], <https://iif.dl.itc.u-tokyo.ac.jp/repo/s/uparl/document/f30080b8-e0fc-448f-bf27-b3c6f1a9323b>, (参照 2018-10-31).
- [13]. 宇野雪村：法帖事典，雄山閣（1984）。
- [14]. 北本朝展，堀井洋，堀井美里，鈴木親彦，山本和明：時系列史料の機械分擔構造化：古典籍『武鑑』を参照する江戸情報基盤の構築に向けて。人文科学とコンピュータシンポジウム じんもんこん 2017, pp.273-280（2017）。