

# 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

中村覚<sup>†1</sup> 成田健太郎<sup>†1</sup> 永井正勝<sup>†1</sup> 富澤かな<sup>†1</sup>

**概要**：東京大学附属図書館 U-PARL（アジア研究図書館上廣倫理財団寄付研究部門）は、本学所蔵資料から選定した漢籍・碑帖拓本（ひじょうたくほん）の資料画像を flickr 上で公開している。本研究では、これら資料の研究利用に向けた要件定義と、Omeka と IIIF を用いて試作したデジタルアーカイブと研究環境について述べる。さらに、集帖を対象とした作品メタデータ（作品タイトルや著者名）の付与、およびテキストアノテーションの付与を行うケーススタディを通じ、構築した研究環境の有用性について検討する。

**キーワード**：漢籍、碑帖拓本、Omeka、IIIF、デジタルアーカイブ

## Prototyping of Digital Library and Application to Philological Studies of Rare Chinese Books and Calligraphic Models in U-PARL

SATORU NAKAMURA<sup>†1</sup> KENTARO NARITA<sup>†1</sup> MASAKATSU NAGAI<sup>†1</sup>  
KANA TOMIZAWA<sup>†1</sup>

### 1. はじめに

東京大学附属図書館 U-PARL（アジア研究図書館上廣倫理財団寄付研究部門）は、本学所蔵資料から選定した漢籍・碑帖拓本（ひじょうたくほん）の資料画像を flickr 上で公開している。本稿では、2 章でこれら資料公開の現状を概観したのち、3 章において研究利用に向けた要件定義を行う。この要件定義に基づき、試作したデジタルアーカイブと研究環境を 4 章で述べる。このシステム構築においては、画像を公開し共有するための国際的な規格である IIIF（International Image Interoperability Framework）[1][2]を採用し、実装にはオープンソースの Web パブリッシングのプラットフォームである「Omeka[3]」を用いる。5 章では要件定義に基づくシステムの機能検証を目的としたケーススタディを行い、6 章で構築した研究環境の有用性について考察する。

### 2. U-PARL における漢籍・碑帖拓本資料の flickr 公開

#### 2.1 資料のデジタル化の条件と方針

U-PARL[4]は、東京大学総合図書館の大幅な拡充計画「新図書館計画」の一環である、「アジア研究図書館」の設立と運営に参加している寄付研究部門である。本部門の課題は多岐にわたるが、アジア資料の適切なデジタル化と公開は特に重視されている。種々の方法を検討した結果、U-PARL

においては flickr を利用して画像を公開することとし、現在のところ漢籍・碑帖拓本資料の公開を行なっている[5][6]。

U-PARL は、寄付に依拠する時限プロジェクトであり、また、人文学・社会科学研究者が主体となっているため、独自のアーカイブの構築と持続は条件的に難しい。加えて、東京大学では大学全体のデジタルアーカイブ基盤構築事業が進んでおり、ゆくゆくはより広域の横断的ネットワークとの連携も想定される。そのため U-PARL では、簡易で低コストの最小限の構成でありながらも、「見つかる＝ウェブで検索される」、「使える＝フルサイズ JPEG とメタデータをダウンロードして、明確なライセンス表示のもとに自由に使うことができる」、「持続する＝維持の難しいサーバやシステムを独自に持つことはせず、上位のアーカイブ基盤が成立したらそちらに連携していく」アーカイブにするという方針を立て、これに基づき、flickr での資料公開という方法を選んだ。

#### 2.2 基本単位の設定と flickr の構造

U-PARL でデジタル化している資料は、製本された図書を基本とする旧来の図書館の枠組み内に位置づけられている。このため、資料の基本単位も、個々のデジタル画像レベルではなく、書誌に置くこととした。ここでいう書誌とは NACSIS-CAT などの総合目録に登録する際の基本単位であるが、以下本稿ではこの書誌単位を「資料」と呼ぶことにする。メタデータを付与する基本単位も、画像ではなく、この「資料」単位で行った。本研究で「メタデータ」

<sup>†1</sup> 東京大学  
The University of Tokyo.

といった場合、特にことわりのない限り、この資料単位のメタデータを指す。

資料という単位の設定で注意すべきは、1 資料が 1 冊になるとは限らず、複数冊で 1 資料になる場合があることである。これを踏まえ、flickr では以下の階層を設定して資料と画像を収めた。つまり、album とその上位の collection という階層的なフォルダーが flickr に用意されていることを利用し、図 1 に示すように、資料単位を collection で、冊単位を album で整理することとした。

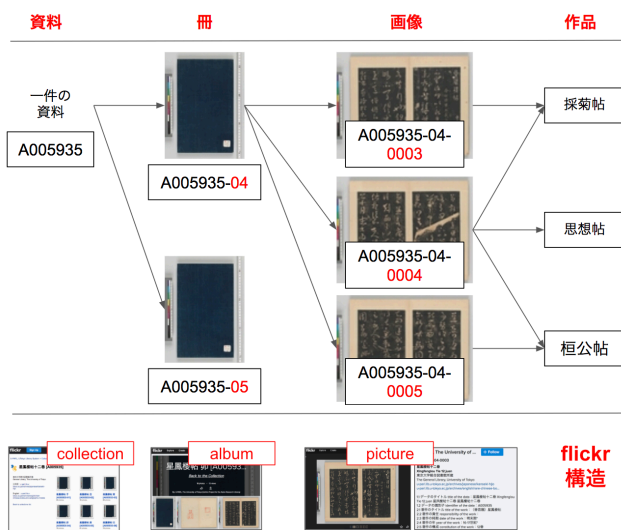


図 1 資料の基本単位と flickr の構造  
 Figure 1 Units of books and structure of flickr.

### 2.3 公開資料の概略とライセンス表示

U-PARL では現在、表 1 に示す漢籍・碑帖拓本 11 資料を公開している。

表 1 公開資料の概略  
 Table 1 Overview of contents.

ID	名称	概要
A005935	星鳳楼帖十二卷	中国で刊行された法帖(集帖)。東京大学総合図書館蔵。
A005936	偽絳帖十二卷	中国で刊行された法帖(集帖)。東京大学総合図書館蔵。
A005939	宣和秘閣帖残一卷	中国で刊行された法帖(集帖)。東京大学総合図書館蔵。
A006096	戲鴻堂法書十六卷	明・董其昌輯。中国で刊行された法帖(集帖)。東京大学総合図書館蔵。

A006099	泉州本淳化閣帖十卷闕四卷	北宋・王著輯。中国で刊行された法帖(集帖)。東京大学総合図書館蔵。
A006100	停雲館帖十二卷	明・文徵明輯。中国で刊行された法帖(集帖)。東京大学総合図書館蔵。
A006107	偽絳帖残一卷	中国で刊行された法帖(集帖)。東京大学総合図書館蔵。
ASIA0001	北路紀略四卷	朝鮮・鄭允容著。漢籍写本。U-PARL 購入。
ASIA0002	古今歴代法帖	朝鮮・朴文会輯。朝鮮で刊行された版刻法帖(集帖)。U-PARL 購入。
ASIA0003	朴彭年草書千字文	朝鮮で刊行された版刻法帖(単帖)。U-PARL 購入。
ASIA0004	金麟厚草書千字文	朝鮮で刊行された版刻法帖(単帖)。U-PARL 購入。

漢籍とは、本稿では主に前近代の東アジアで漢文を使って撰述され、刊行または書写された書籍をいう。碑帖拓本資料とは、碑、銘、墓誌などの石刻から採拓した拓本と、法帖原石から採拓した拓本の総称であるが、現在のところ公開資料に前者は含まれていない。また本稿で法帖とは、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おり、石刻拓本も、当初より石に刻された点を除けばこれに類するものである。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収めた専帖(せんじょう)、複数の筆者の作品を収めた集帖(しゅうじょう)あるいは叢帖(そうじょう)の区別がある。

今回のデジタル公開で対象とした 11 資料は、漢籍 1 資料、法帖 10 資料より成り、法帖のうち 3 資料は、法帖において一般的な拓本ではなく、木版印刷によって製作された版刻法帖である。また法帖 10 資料のうち、8 資料が集帖、2 資料が単帖である。

以上すべてのコンテンツは、制作時期などから合理的に著作権保護期間の満了が推定できるものとして、「クリエイティブ・コモンズ 表示-非営利-継承 4.0 国際 ライセンス (CC BY-NC-SA)」の条件で提供している。

## 2.4 漢籍・碑帖拓本資料の統合メタデータ

漢籍にせよ碑帖拓本資料にせよ、古典籍の個別資料は、著作以下多くの責任 (responsibility) を経たうえていま私たちの目の前に存在する。一つの著作は異なる責任を経るごとに異なる属性を獲得して個別資料となっているのであり、その来歴を詳らかにすることは人文学, なかなく文献学, 書誌学の重要な課題となる。

最も一般的な木版印刷の漢籍でいえば、少なくともテキストの撰述 (著作) から、版の作成 (刻)、版から紙への印刷、そして装訂まで幾層にもわたる責任がある。法帖 (集帖) の場合も、書跡の選定 (輯)、石材への転写・鐫刻、原石から紙への採拓、装訂などがある。

現行の各種目録規則やメタデータ規格は、このような責任の重層性を記述し分けることを想定していないため、U-PARL では漢籍・碑帖拓本資料に適したメタデータ項目を独自に考案し、上述の各階層の責任をメタデータに記述することにした (表 2) [7].

表 2 漢籍・碑帖拓本統合メタデータ項目

Table 2 Metadata categories for rare Chinese books and calligraphic models.

項目 No.	項目名 (日本語)	項目名 (英語)
1.1	データのタイトル	title of the data
1.2	データの識別子	identifier of the data
2.1	著作のタイトル	title of the work
2.2	著作の責任	responsibility of the work
2.3	著作の時期	date of the work
2.4	著作の年	year of the work
2.5	著作の構成	constitution of the work
3.1	バージョンの責任	responsibility of the version
3.2	バージョンの成立時期	date of the version
3.3	バージョンの成立年	year of the version
3.4	バージョンの成立場所	place of the version
3.5	バージョンの技法	technique of the version
3.6	バージョンの大きさ (縦×横)	dimensions of the version (height x width)
3.7	バージョンの行字数	lineage of the version
4.1	資料化の責任	responsibility of the materialization
4.2	資料化の時期	date of the materialization

4.3	資料化の年	year of the materialization
4.4	資料化の場所	place of the materialization
4.5	資料化の技法	technique of the materialization
5.1	外題	external title
5.2	数量	quantity
5.3	装訂	bibliopeggy
5.4	大きさ (縦×横)	dimensions (height x width)
5.5	付加的記載	additional description
5.6	印記	seal
5.7	出所	provenance
5.8	状態	condition
6.1	請求記号	call number
6.2	登録番号	registration number
6.3	原簿番号	ledger number
7	参考文献	reference

このメタデータ項目では、漢籍・碑帖拓本資料の責任を 2.著作, 3.バージョン, 4.資料化, 5.個別資料の 4 階層に分け、さらにデジタル撮影によって生成された 1.データの階層を加えている。これにより、資料の著作レベルから個別資料レベルに至る属性を整理して記述することができ、そのことは漢籍・碑帖拓本資料を用いた人文学研究の目的に適うものと認められる。

ただし、メタデータは資料単位で付与しており、資料中の各部分の属性は記述対象としていない。漢籍・碑帖拓本資料の研究利用にさらなる利便を提供するためには、次章に述べるような要件をクリアする必要がある。

## 3. 漢籍・碑帖拓本資料の研究利用に向けた要件定義

### 3.1 様々なデジタルアーカイブで公開されている資料への横断的なアクセス

一般に古典籍は、同一著作であっても個別資料ごとに異なる属性を有し、その比較検討が研究上重要な課題となる。古典籍のデジタル化・オープン化が進展したときに望まれるのは、著作レベルで相互に関連する複数の個別資料に横断的にアクセスできることであり、そのためには、著作単位の典拠データが準備され、異なる個別資料が同一の典拠データに結びつけられている必要がある。

NACSIS-CAT では、参加する多数の所蔵機関が書誌データを共有する運用がなされているが、古典籍についてはそ

のような運用がなされず、通常個別資料単位で書誌データが作成されている。というのも、新書の場合は、著作レベルないし版レベルの異同の判断に必要な情報を個別資料現物から得て書誌データ上に記述する方式が確立した十分に共有されていて、他の所蔵機関が作成した書誌データに自機関所蔵の個別資料を紐付けることが容易にできるが、古典籍の場合、個別資料から必要な情報を獲得する方法が広く共有されているわけではなく、他機関の作成した書誌データと自機関所蔵の個別資料とを照らし合わせても、著作レベルの異同は必ずしも明らかにできない。たとえばタイトルが異なっても同一著作である可能性、反対にタイトルが同じでも別個の著作である可能性があり、その判断には専門知識が必要となる。

ある程度の分量の漢籍を所蔵している収蔵機関は、多く漢籍目録を作成しており、その記載には著作レベルの同定に必要な情報が抽出されていて、漢籍の書誌データ共有を実現するうえで参照価値が高い。たとえば中国古代の思想家である揚雄（ようゆう、BC53-AD18）の著作『法言（ほうげん）』は、あるいは著者の尊称を冠して『揚子法言（ようしほうげん）』とも呼ばれる。『東京大学東洋文化研究所漢籍分類目録』は、『法言』と『揚子法言』のタイトルの異同を無視して計 21 資料を一か所にまとめて掲載している [8]。さらに漢籍目録では一般に、タイトルの直後に書誌的巻数を記す習慣があり、『法言』ないし『揚子法言』の場合、主として十巻と記されるものと十三巻と記されるものがある。この書誌的巻数は、個別資料に固有の物理的分量ではなく、著作に固有の属性であって、巻数が異なれば別個の著作と見なされることがある。

さらに撰・注等の責任を記す項には、十巻本の場合「漢揚雄 撰」のみ記すものと、これに加えて「晋 李軌 唐 柳宗元 注」をも記すものがあり、他方十三巻本の場合「漢揚雄 撰」「晋 李軌 注」が記されている。つまり、同じ『法言』ないし『揚子法言』でも、10 巻から成り揚雄の本文のみ収めるもの、13 巻から成り揚雄の本文に加えて李軌（りき、4 世紀頃）の注をも収めるもの、そして 10 巻から成り揚雄の本文、李軌の注、さらに柳宗元（りゅうそうげん、773-819）の注を収めるものがあり、これらは構成要素を異にするので、やはり著作レベルの異同と見なされる。

このように、漢籍の著作レベルの異同を認定するためには、著作に固有の構成要素を把握する必要があり、書誌データには構成要素に関する情報が適切に記述されている必要がある。

しかしながら古典籍の場合、個別資料に残欠や改変があるなどの理由によって、最初に書誌データを作成する機関において著作レベルの典拠データを完成できるとは限らず、またそれが十全であるかどうかを直ちに判断することもで

きない。したがって、最初に書誌データを作成してしまってからそれを共有する NACSIS-CAT のような形ではなく、最初は個別資料レベルの情報を記述していった、後にそれらを統合して著作レベルの典拠データを獲得するという流れが望ましいと思われる。

### 3.2 「作品メタデータ」と「テキストアノテーション」

以上のように、著作レベルの典拠データを獲得するためには、まず個別資料に即してその構成要素を記述する必要がある。具体的には、章、作品など構成要素ごとのタイトル、さらに著者名が記述されるのがよく、これらを画像データにメタデータとして付与する手順が考えられる。本研究ではこれを「作品メタデータ」と呼ぶ。たとえば上述の『法言』の巻 1 には章タイトル「学行（がっこう）」を、蕭統（しょうとう、501-531）によって編まれた詞華集『文選（もんぜん）』ならば巻 1 に作品タイトル「両都賦（りょうとふ）」と著者名「班固（はんこ、32-92）」をそれぞれメタデータとして付与する。上述のように、漢籍目録においては著作レベルの異同を明らかにするために書誌的巻数を記述することが一般に行われるが、構成要素ごとのタイトル・著者名を記述することは煩雑を招くので通常なされない。すなわち、著作の同定に漢籍目録よりも精度の高い情報を提供することができ、ここにメタデータ方式の大きな利点が認められるのである。

碑帖拓本資料、なかんずく法帖にも、多数の筆者による多数の作品から成る集帖があり、構成要素ごとのタイトル・著者名をメタデータとして付与することはやはり著作の同定に大きく寄与しうる。ただし、集帖に収録される個別の作品は、大部分の漢籍と異なって資料中に作品タイトルを明示しないことが多く、その場合テキスト中の連続する数文字を随意選択して便宜上のタイトルとする習慣がある。そのような作品タイトルは、300 種以上の集帖・専帖を調査して構成要素を書き出した容庚（ようこう、1894-1983）の著作『叢帖目（そうじょうもく）』[9]に最も多く集積されており、これを一種の典拠データとして利用することができる。ただし、作品タイトルはテキスト中から任意に取られるため異称が多く、『叢帖目』にも同一作品が場所によって別のタイトルで採録されている瑕疵がある。また、複数の異なる作品から偶然同一の文字列が選択されることも起こりうるため、例えば王羲之（おうぎし、303-361）には「初月帖」というタイトルを持つ作品が複数ある。したがって、作品の同定においては作品タイトルに過度に依存しないよう注意しなければならない。著者名についても、資料から適切に著者を同定して記述するためには一定の専門知識が必要である。

そこで次には、各作品の画像の文字列部分に、アノテー

ションとしてテキストデータを付与する手順が考えられる。本研究ではこのテキストデータを「テキストアノテーション」と呼ぶ。これにより、作品の同定に豊富な情報を提供することができる。各作品について作品メタデータとテキストアノテーションが蓄積されれば、それらを統合して作品ごとの典拠データを作成する段階も視野に入ってくるだろう。そうすれば典拠データを介して異なる集帖に刻入された同一作品を比較検討することが容易になり、法帖研究に大きな利便をもたらすことができる。古典籍のなかでも特に法帖は、比較的少数の共通する作品が多数の集帖に刻入されるので、このようなモデルは十分な実現可能性を有し、研究の有効な方策として成長が期待できる。

#### 4. Omeka を用いたデジタルアーカイブと研究環境の構築

本章では前節の要件に基づく研究環境の構築を行う。具体的には図 2 に示すように、flickr 上で公開されている画像資料を IIIF に準拠して公開するデジタルアーカイブと、研究者がメタデータやアノテーションを付与可能な研究環境を構築する。これらのシステム実装には、オープンソースの Web パブリッシングのプラットフォームである「Omeka」を用いる。Omeka には、コンテンツ提供機関向けの「Omeka S」と、コンテンツ利用者向けの「Omeka Classic」がある。本研究では、前者を用いてデジタルアーカイブを構築し、後者を用いて研究環境を構築する。

以下では、各々のシステムについて述べる。



図 2 Omeka を用いたデジタルアーカイブと研究環境  
 Figure 2 System overview with Omeka.

##### 4.1 「Omeka S」を用いたデジタルアーカイブ構築

まず、「Omeka S」を用いたデジタルアーカイブの構築について述べる。flickr で公開済みの画像を Omeka S を用いて再公開する理由としては、後述する研究環境が IIIF に準拠した画像を必要するためである。

実装には、Omeka S のモジュール (プラグイン) の一つである「Omeka-S-module-iiifServer[10]」を用いることにより、Omeka S に登録したメディア (画像) を IIIF Image API

を用いて公開することができる。さらに、モジュール「Omeka-S-module-UniversalViewer[11]」を用いることにより、Omeka S に登録したメタデータに基づいて、IIIF マニフェストを自動的に作成することができる。

図 3 にこれらモジュールを用いて試作したデジタルアーカイブの画面例を示す。図左部は検索画面であり、2.4 で述べた漢籍・碑帖拓本資料のメタデータに基づいて資料を検索することができる。図中央部は画像の閲覧画面であり、Universal Viewer を用いて閲覧する。図右部は資料のメタデータを表示する画面である。本研究では、2.3 のコンテンツの概略で述べた 11 資料、計 67 冊、計 1,889 画像を Omeka S に登録した。



図 3 Omeka S を用いたデジタルアーカイブの画面例  
 Figure 3 Interface of digital archives with Omeka S.

##### 4.2 「Omeka Classic」を用いた研究環境の構築

次に「Omeka Classic」を用いて、研究者がメタデータやアノテーションを付与・管理することが可能な研究環境の構築について述べる。Omeka Classic のプラグインの一つである「IIIF Toolkit with Mirador[12]」を用いることにより、ウェブ上の IIIF マニフェストを用いて画像をインポートし、それらにアノテーションを付与することが可能となる。この機能を用いて、前節で構築したデジタルアーカイブが提供する IIIF マニフェストをインポートすることにより、研究者によるアノテーションの付与が可能な研究環境を構築した。この研究環境については、次章のケーススタディで詳述する。

#### 5. ケーススタディ

3 章の漢籍・碑帖拓本資料の研究利用に向けた要件定義において、著作レベルの典拠データを獲得することの必要性を述べた。具体的には研究利用において、資料単位ではなく、作品単位での「作品メタデータ」および「テキストアノテーション」が有益な情報となる。また「典拠データを介して集帖に刻入された同一作品を比較検討する」ことの必要性について述べた。

本ケーススタディでは、構築した研究環境が、これらの

要件を満たすことができるかを検証する。まずは集帖に関する先行研究を当座の典拠データとして利用し、作品タイトルと筆者の情報を画像レベルに反映して、集帖中の各作品に絞ってデータにアクセスできる環境の構築可能性を示す。さらにテキストアノテーションをも加えることで、典拠データと個別資料に即したテキスト情報とを、相互に参照・検証しつつ整備することが可能になるという展望を示す。

### 5.1 作品メタデータの付与

「王羲之」の作品ごとに「採菊帖」や「思想帖」等の作品タイトルをメタデータとして Omeka Classic に登録する。登録した作品のメタデータの例を表 3 に示す。

表 3 登録した作品メタデータの例

Table 3 Example of metadata about titles and authors.

資料 ID	資料名	冊次	作品タイトル	筆者	画像番号(範囲)
A005935	星鳳楼帖	04	採菊帖	王羲之	p. 0003 - 0004
A005935	星鳳楼帖	04	思想帖	王羲之	p. 0004
A005935	星鳳楼帖	04	桓公帖	王羲之	p. 0004 - 0005

「資料 ID」「資料名」は各集帖に資料単位で与えられるメタデータである。一方、「作品タイトル」「筆者」は集帖中の作品単位のメタデータである。「画像番号」列に示すように、画像に作品メタデータ（作品タイトルと筆者）を付与した。図 4 には Omeka Classic を用いたメタデータの登録画面と、作品単位での検索画面を示す。作品タイトル「採菊帖」や筆者「王羲之」等の作品メタデータを用いた絞り込みができ、従来の資料単位よりも小さな粒度での検索が可能な環境を構築した。



図 4 作品単位でのメタデータ付与と検索画面例

Figure 4 Interfaces to accumulate and search metadata.

### 5.2 集帖に刻入された同一作品の比較検討

さらに、図 5 に示すように、新日本古典籍総合データベース[13]で公開されている集帖「歴代帝王法帖[a] [14]」に含まれる作品「採菊帖」と、U-PARL が公開する「採菊帖」の比較を IIIF 対応ビューワの一つである「Mirador[15]」を用いて行なった。これは、3 章で研究利用の一例として挙げた同一作品の比較検討を可能とする機能である。このように、異なるシステムで公開されている画像を共通のビューワで比較可能な点は、IIIF を採用した本研究環境の利点を示す一例である。

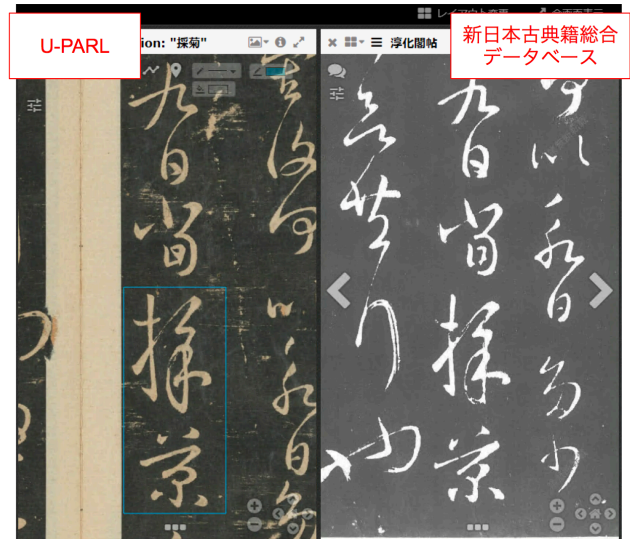


図 5 Mirador を用いた作品の比較例

Figure 5 Comparison of works with Mirador.

a) 当該データベースではタイトルを「歴代帝王法帖」としているが、これは同資料の一部の構成要素に対する名称であり、著作のタイトルは「淳

化閣帖」とすべきである。

### 5.3 テキストアノテーションの付与

上記の作品メタデータの登録に加え、初歩的な実践として、作品識別の手がかりとして用いられている文字列（例えば「採菊帖」の「採菊」）をテキストアノテーションとして付与した。これにより、典拠データに対して個別資料の画像がデータの根拠として結びつき、紙媒体の先行研究よりも一歩進んだ典拠データの形を示すことができた（図6）。

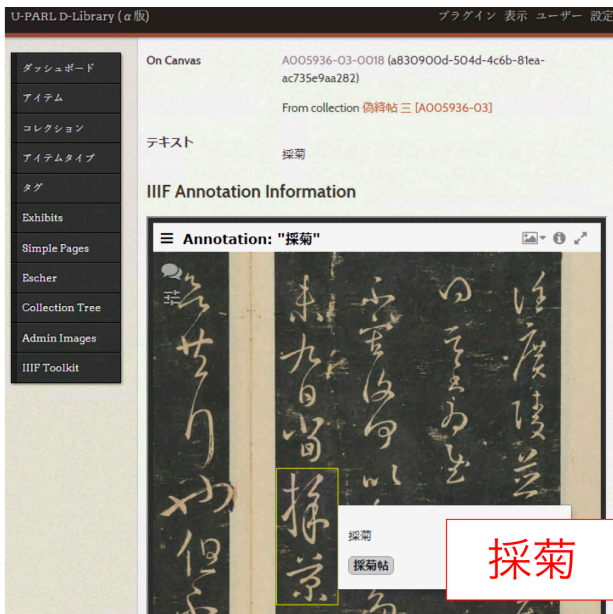


図6 テキストアノテーションの付与例  
 Figure 6 Example of Text annotation.

さらに、すでに述べたように、集帖に収められる各作品のタイトルには異称を持つものがあり、例えば王羲之の「思想帖」には「不復見君帖」の異称がある。このような場合、2つの作品タイトルに取られている文字列をどちらもアノテーションしておけば、一方の作品タイトルからアクセスした利用者ないし利用機会に、もう一方の異称の存在を知らせる機能を持たせることができる（図7）。

すなわち、典拠データを整備していく過程のなかで、作品タイトルの異称を名寄せする作業に有益な情報を提供し、なおかつそれが個別資料の画像とつねに結びついている有利な環境を実現することができた。



図7 異称のテキストアノテーション例  
 Figure 7 Text annotations of variant titles.

## 6. 考察

### 6.1 人文学研究から見た意義と課題

人文学研究、なかんずく本研究に大きく関わる文献学・書誌学研究では、従来著作、バージョン、個別資料、テキスト、文字に対する研究が目録、解題、書誌、校訂などの異なるレベルで個別に行われてきた。本研究は、資料のメタデータから、作品メタデータ、テキストアノテーションに至るまで、人文学研究に有用な各レベルの情報をシームレスに集約するモデルを示している。このことは、知的所産として追究すべき著作のレベルと、具体的材料として依拠すべき個別資料のレベルを逐次行きつ戻りつして考察を積み上げなければならない人文学研究の営みに有利な状況を提供する。5.3で示したようなテキストアノテーションは、いわば資料のごく一部分にのみ翻刻を施した状態であり、そのような状態は資料に対する研究の過程において必ず発生するが、紙媒体を前提とする研究成果公開の枠組みにおいては、全テキストの翻刻まで作業を進めてから公開することが求められてきた。本研究では、作業途中の段階でも成果を逐次公開し、それを共有してさらなる向上に資することができるという、人文学研究の新たなモデルにかなう実践といえるだろう。

一方で、アノテーションの実践には課題もある。例えば5.3で述べた集帖中の作品タイトルのテキストアノテーションに関して、作品タイトルの根拠となる文字列が2行にまたがっている場合、テキストアノテーションが2つの範囲に分かれてしまう。さらにその2行は2画像にまたがる可能性もあり、そのような場合にはどのようなアノテーション方法が適当か、なお検討が必要である。さらに、将来

的にアノテーションの対象を全文字列に拡大すると、同一作品のテキストを複数画像にまたがってアノテーションするケースは非常に多くなる。このことが、作品ごとに情報を整備して作品ごとの典拠データに反映していく過程にどのように影響するかもまた、検討すべき課題となるだろう。

また、CiNiiなどの総合目録検索システムでは、日本の常用漢字体・旧字体や中国の簡体字・繁体字などの異体字相互の対応関係がデータとして準備されており、字体の違いに起因する検索漏れを排除している。本研究の趣旨に照らせば、Omekaにプラグインとしてそのような機能が導入できると望ましい。さらに、個別資料の状態によって文字が残余したり判別できなかったケースにも対応できるよう、テキスト検索にあいまい検索の機能も付け加えられるとなお好ましい。

## 6.2 オープンソースソフトウェアと国際規格の利用

本研究で構築した研究環境が提供する「メタデータ登録」や「アノテーション付与」「画像比較」等の機能群は、新しい技術や機能ではない。一方、本研究ではオープンソースソフトウェアである Omeka を用いて環境構築を行うことにより、第三者による再現が容易な研究環境を提案した点に有用性があると考えられる。また、本環境で取り扱う画像やアノテーション等の情報が IIF 準拠で格納されることは、データの相互運用性が高いという点において、本研究環境の有用性を示す一つの特徴であると考えられる。

## 6.3 公開の実現性と研究深化の両立

U-PARL が flickr に公開した資料は、書誌を基本単位とする図書館の論理に即し、1冊または複数冊から成る資料単位でメタデータを付与している。これはライブラリアンが主体となる図書館資料のデジタル化にあたっては取り組みやすく、既存の書誌や OPAC との連動もしやすい現実的な方策と言えよう。一方で、研究利用にあたっては、5章で示したように、より深く細かいデータが求められる。資料のデジタル化の効率や持続性を考えた場合、資料を所蔵しデジタル化する側が深いデータを提供できるとは限らない。それゆえ、デジタル資料の公開においては、資料が見つかり、適切なライセンス表示のもとに自由に使い、その後も持続する状況を整えておけば、研究に即した多様な作業が後に発展していくことが可能となるはずである。本研究が提示したモデルは、図書館主体の書誌単位のデジタル化を、研究者主体の深いデータ化に結びつけることを可能にするものであると言える。

## 7. 結論

本研究では、東京大学附属図書館 U-PARL (アジア研究図書館上廣倫理財団寄付研究部門) が公開している漢籍・碑帖拓本資料を対象として、これら資料の研究利用に向けた要件定義を行った。その結果として、著作レベルの典拠データの獲得に向けた作品メタデータ (作品タイトル、著者名)、テキストアノテーションの重要性を述べ、この要件定義に基づく研究環境を構築した。具体的には、Omeka と IIF に関連する各種プラグインを用いて、IIF に準拠して画像を公開するデジタルアーカイブと、研究者がメタデータやアノテーションを付与することが可能な環境を構築した。このシステムを用いて作品メタデータ (作品タイトル、著者名) や、テキストアノテーションを付与することで、作品の同定に必要な情報を蓄積し、絞り込み等の資料検索に活用するモデルを示すことができた。

本研究では、構築したシステムの機能検証に留まったが、今後は具体的な課題を定め、実際の研究利用におけるシステムの有用性を検証する。

## 参考文献

- [1] “IIF | International Image Interoperability Framework”. <http://iiif.io/>, (参照 2017-12-21).
- [2] 永崎研宣. デジタル文化資料の国際化に向けて: IIF と TEI. 情報の科学と技術, 2017, vol. 67, no.2, pp. 61-66.
- [3] “Omeka”. <https://omeka.org/>, (参照 2017-12-13).
- [4] “東京大学附属図書館アジア研究図書館上廣倫理財団寄付研究部門 (U-PARL)”. <http://u-parl.lib.u-tokyo.ac.jp/>, (参照 2017-12-24).
- [5] “U-PARL 【DIGITAL LIBRARY】漢籍・碑帖拓本資料”. <http://u-parl.lib.u-tokyo.ac.jp/archives/japanese/kanseki-hijo/>, (参照 2017-12-24).
- [6] “U-PARL, UTokyo Library System's collections on Flickr”. <https://www.flickr.com/photos/asianresearchlibrary/collections/>, (参照 2017-12-24).
- [7] 成田健太郎. 碑帖拓本資料のデジタル公開における書誌記述の実践. 書学書道史研究, 2017, no. 27, pp. 29-41.
- [8] 東京大学東洋文化研究所. 東京大学東洋文化研究所漢籍分類目録. 汲古書院, 1981, pp.445-446.
- [9] 容庚. 叢帖目. 中華書局香港分局, 1980-1986.
- [10] “Daniel-KM/Omeka-S-module-IiifServer”. <https://github.com/Daniel-KM/Omeka-S-module-IiifServer>, (参照 2017-12-13).
- [11] “Daniel-KM/Omeka-S-module-UniversalViewer”. <https://github.com/Daniel-KM/Omeka-S-module-UniversalViewer>, (参照 2017-12-13).
- [12] “utlib/IiifItems”. <https://github.com/utlib/IiifItems>, (参照 2017-12-13).
- [13] “新日本古典籍総合データベース”. <http://kotenseki.nijl.ac.jp/>, (参照 2017-12-21).
- [14] 歴代帝王法帖, 書誌 ID:100065344, DOI: 10.20730/100065344
- [15] “Mirador”. <http://projectmirador.org/>, (参照 2017-12-25).