

京都大学貴重資料デジタルアーカイブ

現在までの取り組み

西岡 千文 (にしおか ちふみ)

✳ 1. はじめに

京都大学附属図書館では、過去 20 年以上にわたり、所蔵している貴重資料の電子化・公開を進めてきた。近年、京都大学オープンアクセス推進事業^[1]等により、人文社会科学系研究基盤の強化を目的とした古典籍等の貴重資料の電子化・公開をさらに積極的に進めている。これらを背景として、京都大学貴重資料デジタルアーカイブ^[2]は、2017 年 9 月 7 日に試験公開され、2017 年 12 月 1 日に正式公開となった。2017 年 12 月 1 日現在、3,719 タイトル、394,069 画像を公開している。デジタルアーカイブは、画像の国際的な相互利用の促進を目的とした枠組みである IIIF (International Image Interoperability Framework)^[3]に準拠している。

本稿は、京都大学貴重資料デジタルアーカイブを紹介することを目的とする。次章で IIIF について短く紹介し、3 章で京都大学貴重資料デジタルアーカイブについて基本的な機能とシステム構成について述べる。4 章で筆者が現在取り組んでいる各画像の利用状況の可視化について述べ、5 章をむすびとする。

✳ 2. IIIF 概要

これまでに数多くのデジタルアーカイブが公開されてきたが、画像が機関によって異なる方法で提供され、共有や再利用が難しかった。このよ

うな状況を改善するために、IIIF の取り組みが開始された。画像の国際的な相互利用を実現するために、共通の API と API に対応した画像サーバ、ビューワ等ソフトウェアの開発をコミュニティで行っている。2018 年 1 月現在、4 種の API が公開されているが、コアとなる API は Image API と Presentation API である。京都大学貴重資料デジタルアーカイブはこれら 2 種の API を使用している。以下、Image API と Presentation API について短く紹介する。

● 2.1. Image API

Image API は以下の URI 構文^[4]「{scheme}://{server}/{prefix}/{identifier}/{region}/{size}/{rotation}/{quality}.{format}」によって画像の呼び出しを行う。この URI 構文により、表示箇所 (region)、サイズ (size)、角度 (rotation) 等を指定して画像を呼び出すことが可能である。サムネイルとして使用する場合は、サイズを小さく指定して画像を呼び出す。また、IIIF 対応ビューワでユーザが画像を拡大・縮小する度に、画像の箇所・サイズを変更して画像の呼び出しが行われている。

● 2.2. Presentation API

Presentation API は、書誌名、権利情報等のメタデータや、資料の中でどのような順序で画像が登場するかといった情報の記述形式を定義する。Presentation API では、一つの資料がマニフェスト

という要素に対応している。各資料のマニフェストにはその資料を表示するために必要な情報が格納され、JSON 形式のファイルに記述される。マ

ニフェストには以下の情報^[5]が含まれる。

- JSON が IIIF マニフェストファイルであることを示す情報
- 資料全体に関する情報 (書誌名、権利情報等)
- 資料を構成する部分についての情報

資料を構成する部分に関する情報について述べる。マニフェストにおいて資料の各項はキャンバスという要素に対応する。各キャンバスは、キャンバスに描画する画像・注釈 (翻刻・翻訳等) についての情報をもつ。このデータ表現によって、同項に属していたが散逸してしまった資料を、デジタル空間上で一つの項に表示することが可能である^[6]。また、キャンバスに画像を描画しないことにより、欠損した項を表現することが可能である。マニフェストは、キャンバスを順番に並べたシーケンスを含むことで、資料の項の順序を表現する。

Presentation API が定義する形式に沿って資料に関する情報をマニフェストファイルに記述することで、あらゆる IIIF 対応ビューワで資料を閲覧することが可能となる。

✳ 3. 京都大学貴重資料デジタルアーカイブ

本章では、京都大学貴重資料デジタルアーカイブの基本機能とシステム構成について述べる。

● 3.1. 基本機能

本節では、京都大学貴重資料デジタルアーカイブの基本的な機能をユーザインタフェースとともに紹介する。

■ トップページ (図 1)

トップページには、「お知らせ」等とともに、「ピックアップ」と「コレクション」が表示されている。「ピックアップ」は国宝・重要文化財等、頻繁に利用される貴重資料をカテゴリ別に表示する。「コレクション」では、公開されている全資料を文庫別に表示している。右上に位置するボタンで日本語・英語の言語の切り替えが可能である。

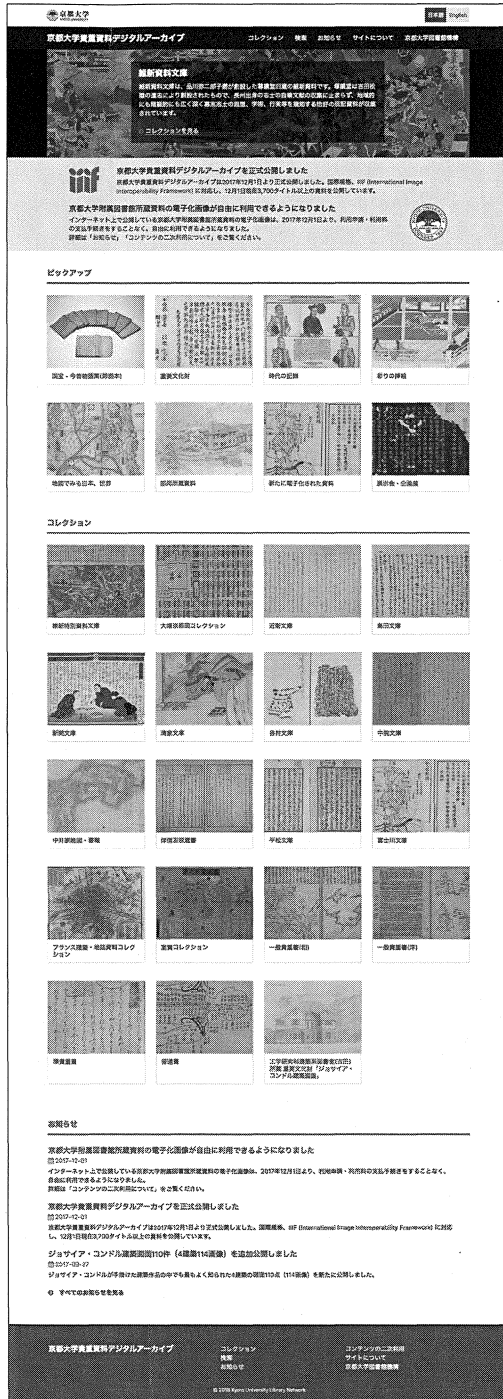


図 1: トップページ

国際的な学術情報流通の促進を目指して、書誌名等のメタデータのローマ字化にも積極的に取り組んでいる。

■ 検索 (図 2)

テキストボックスに検索クエリを入力すると、書誌情報・翻刻テキストの全文検索が行われ、結果が表示される。

■ コレクション情報 (図 3)

各コレクション(文庫)のページでは、コレクションの内容や由来の解説と、コレクションに含まれる資料一覧を表示する。

■ 書誌情報 (図 4)

コレクション情報のページで任意の書誌をクリックすると、レコード ID、書誌名、注記等の書誌情報が表示される。上部には、IIIF 対応の画像ビューワである Universal Viewer が埋め込まれており、画像を閲覧することが可能である。ビューワ下部には、IIIF マニフェスト、Universal Viewer と Mirador のアイコンが配置されている。IIIF マニフェストアイコンは、IIIF マニフェストファイル (JSON 形式) へのリンクが貼られている。このアイコンを Mirador 等の IIIF 対応ビューワにドラッグ・アンド・ドロップすると、書誌情報が読み込まれ閲覧することができる。Universal Viewer

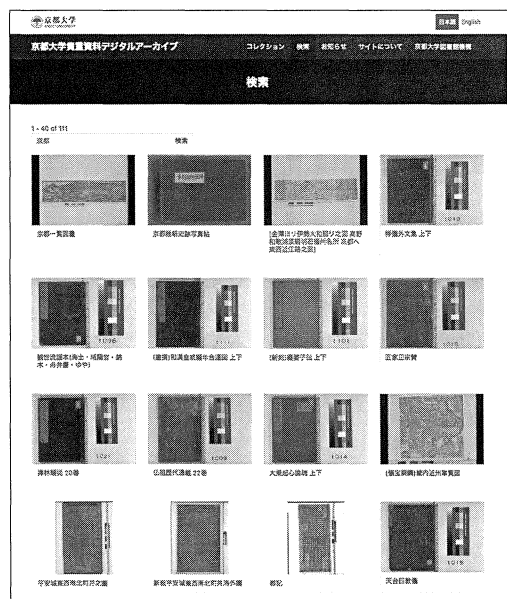


図 2：検索

と Mirador のアイコンをクリックするとそれぞれのビューワが新タブで立ち上がり、画像が表示される。ビューワについては、後述する。

■ ビューワ

IIIF ビューワとして、IIIF のコミュニティで使用頻度が高いかつ開発活動が活発な Universal Viewer^[7] と Mirador^[8] を採用している。特に Mirador は注釈の表示、画面の分割等の機能が豊富である。どちらのビューワも、東洋の古典籍の右から左へのページめくりに対応している。ペー



図 3：コレクション情報



図 4：書誌情報

じめくり方向の指定は、IIIF マニフェストファイルのフィールド `viewingDirection` で行う。

● 3.2. システム構成

デジタルアーカイブシステムは、コンテンツ管理システム（CMS）と画像サーバより構成されている。CMS として、Drupal を採用している。Drupal はオープンソースのモジュラー式フレームワークで、開発の容易さ・公開されている機能の豊富さから、多くのウェブサイトで採用されている。CMS では、貴重資料の書誌情報や解題等のメタデータを格納している。IIIF 画像サーバとして、IIP Image Server を採用している。画像サーバには、画像と IIIF マニフェストファイルが保存されている。画像は Pyramid TIFF 形式で保存されている。IIIF マニフェストファイルは、Drupal に保存されている書誌情報より自動生成される。ビューワで画像が表示される際には、Presentation API と Image API を使用してデータが取得される。また、「ピックアップ」、「コレクション」の書誌一覧が表示される際、サムネイル画像は Image API を使用して取得される。

✳ 4. 利用状況の可視化

IIIF 対応画像ビューワは、Image API によって画像を取得し、表示している。Image API の呼び出しは、ユーザが画像を拡大・縮小する度に発生する。よって、ユーザの画像の利用状況（ユーザが注視した画像の箇所等）を詳細に記録することが可

能である。そこで、Image API の呼び出しログを解析することで、画像の利用状況の可視化を実施した。Python で Image API の呼び出しログを解析し画像の利用状況を表すヒートマップを生成した。

Mirador では、一つのキャンバスへの複数画像の表示、さらに各画像の透過度を調整することが可能である。これらの機能を利用して、貴重資料画像に、その利用状況を示したヒートマップを重ね合わせて表示する（図5）。IIIF が提供する機能を活用することで、簡単に利用状況を表示することが可能である。

✳ 5. おわりに

本稿では、2017 年 12 月 1 日に正式公開された京都大学貴重資料デジタルアーカイブについて紹介した。デジタルアーカイブは、IIIF といった相互利用性の高い枠組みを取り入れていくことで、学術情報の流通促進に貢献する。また、正式公開に合わせ、画像の二次利用が容易になるよう、規定を改正した¹⁹⁾。

今後、数点の翻刻・翻訳が存在する資料については、各翻刻・翻訳の画像における位置情報を取得し、アノテーションとして表示していく。翻刻・翻訳が存在しない資料については、このように利活用が容易な手法で資料を公開することで、その資料に対する研究が進展し、翻刻・翻訳等の情報が増えることを期待している。他にも課題はあるが、研究者にとって有益なデジタルアーカイブへ発展させていきたいと考えている。

謝辞

本稿執筆にあたっては、京都大学附属図書館の大村明美氏、北條風行氏、赤澤久弥氏、富岡達治氏に協力いただいたことを感謝とともに記しておく。

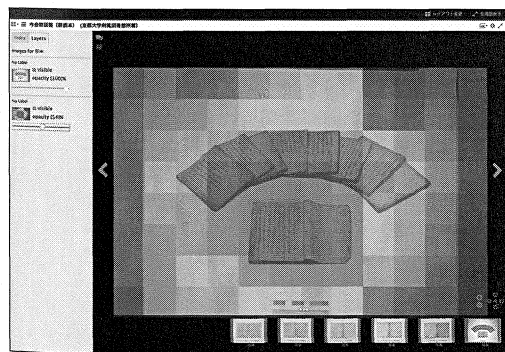


図5：画像の利用状況の可視化

注

- [1] <http://www.kulib.kyoto-u.ac.jp/content0/1373844>
- [2] <https://rmda.kulib.kyoto-u.ac.jp/>
- [3] <http://iiif.io/>
- [4] <http://iiif.io/api/image/2.0/#uri-syntax/>
- [5] 参考資料：<http://www.infocom.co.jp/das/loddiary/iiif/>
- [6] 参考資料：<http://digitalnagasaki.hatenablog.com/entry/2017/07/03/030708>
- [7] <https://universalviewer.io/>
- [8] <https://mirador.stanford.edu/>
- [9] <https://rmda.kulib.kyoto-u.ac.jp/reuse>