

# オープンサイエンスの基盤としての国デコ Image Wall: IIF + Crowd4U の活用事例として

永崎 研宣<sup>†1</sup> 森嶋 厚行<sup>†2</sup> 池田 光雪<sup>†3</sup> 林 亮太<sup>†4</sup> 太田 千尋<sup>†5</sup>

**概要**：日本デジタル・ヒューマニティーズ学会の SIG Transcribe JP では、日本語資料の翻刻をクラウドソーシングで実施する一方で、より効果的なクラウドソーシングを目指し、Crowd4U プロジェクトとの連携によるマイクロクラウドソーシングに取り組んでいる。ここでは、文字の翻刻だけでなく、画像に写真や図表が含まれているかどうかという判定も試行した。そして、後者の成果について、IIF を活用しつつ国デコ Image Wall を構築した。キーワードや前提知識等を必要としない「デジタルアーカイブ」への新たな導線を提供する仕組みは、オープンサイエンスの基盤の一つともなり得ると期待される。

**キーワード**：高精細画像，図書資料，デジタルアーカイブ，メタデータ

## KuniDiCo Image Wall as a Base for Open Science: As an Example of leveraging of IIF and Crowd4U

Kiyonori Nagasaki<sup>†1</sup> Atsuyuki Morishima<sup>†2</sup>  
Kosetsu Ikeda<sup>†3</sup> Ryota Hayashi<sup>†4</sup> Chihiro Ota<sup>†5</sup>

### 1. はじめに

日本デジタル・ヒューマニティーズ学会(a)の分科会 SIG Transcribe JP では、日本語資料の翻刻をクラウドソーシングで実施する取組みとして「翻デジ」サイトを構築・運営する一方で、より効果的なクラウドソーシングを目指し、Crowd4U プロジェクト(b)との連携によるマイクロクラウドソーシングに取り組んでいる。この連携では、文字の翻刻だけでなく、画像に絵や写真・図表が含まれているかどうかという判定も試行した。そして、後者の成果を元にして IIF を活用して、デジタル資料閲覧サイト「国デコ Image Wall」(c)を構築した。

近年再び盛り上がりつつある「デジタルアーカイブ」の議論においては、その利活用の必要性が強調されるようになってきている[1]。しかしながら、キーワードや前提・文脈等についての知識がなければ個々の資料に行き当たることはなかなか容易ではなく、また、検索結果のスニペットやメタデータ、表紙のサムネイル画像等から図書資料の内容に関心を持つことは、基本的な導線ではあるにせよ、これも必ずしも容易なことではない。「国デコ Image Wall」は、そうした情報から入るのではなく、むしろ、図書資料中に

含まれる絵や写真、図表などから入って一つの図書の全体への関心を喚起するという導線を提供するものであり、この仕組みは、デジタルコレクションにおける図書資料の探索に際して、キーワードや前提知識に依拠せずとも可能な手法の一つとして、いわば、デジタルアーカイブ利活用のための「もう一つの導線」の役割を果たし得るものである。類例として、英国図書館が 100 万点以上の絵や図表等の画像を Flickr で公開しているが、これと比較した場合、画像の公開に IIF をも利用しつつ細かな使い勝手を工夫して利便性を高めているという点が、「国デコ Image Wall」の特徴と言えるだろう。ここでは、「国デコ Image Wall」の構築について報告するとともに、画像に絵や写真・図表が含まれていることを確認するための Crowd4U におけるマイクロクラウドソーシングの概要について紹介したい。

### 2. 国デコ Image Wall

国デコ（国立国会図書館デジタルコレクション）は、30 万点以上の図書資料をデジタル化して公開しており、このうち 26 万点超がパブリックドメインである。これらの図書資料の一つ一つに「永続的識別子」が付与され、さらに図書資料の画像一つ一つに画像の順番で番号が付けられ、

†1 一般財団法人人文情報学研究所

International Institute for Digital Humanities

†2 筑波大学知的コミュニティ基盤研究センター

Research Center for Knowledge Communities, University of Tsukuba

†3 千葉大学アカデミック・リンク・センター

Academic Link Center, Chiba University

†4 筑波大学大学院図書館情報メディア研究科

Graduate School of Library, Information and Media Studies, University of

Tsukuba

†5 筑波大学大学院システム情報工学研究科

Graduate School of Systems and Information Engineering, University of

Tsukuba

a <http://www.jadh.org/>

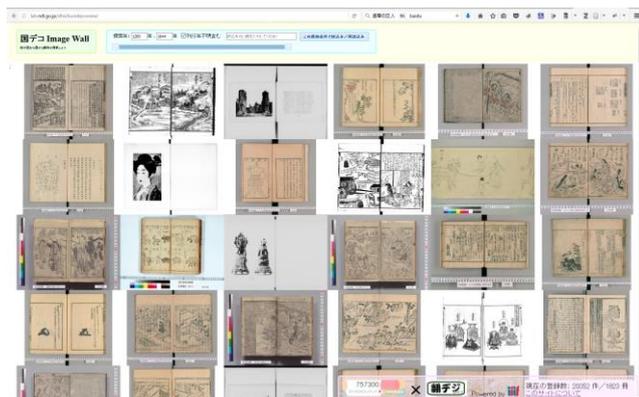
b <http://crowd4u.org/>

c <http://lab.ndl.go.jp/dhii/kunidicoview/>

その番号でアクセスできるようになっていることから、「<http://dl.ndl.go.jp/永続的識別子/画像番号>」という URL (以下、画像 URL) で図書資料の各画像に直接アクセスすることが可能である(d. そこで、絵や写真・図表が含まれている画像 (以下、絵図画像) の画像 URL を収集し、それらを一覧できるようにすることで図書資料にたどり着くという導線を提供することを目指したのが国デコ Image Wall である。

絵図画像の一覧にあたっては、大量のサムネイル画像を一度に表示してしまうことになるが、ここでは Javascript のライブラリである jQuery のプラグイン、Lazy Load<sup>e</sup>を用いることで、ブラウザに表示されるものだけをロードするような仕組みを提供することで、ネットワークやブラウザへの負荷が高くなり過ぎないようにしている。その上で、1 アクセス 300 枚までをランダム表示するようになっている。

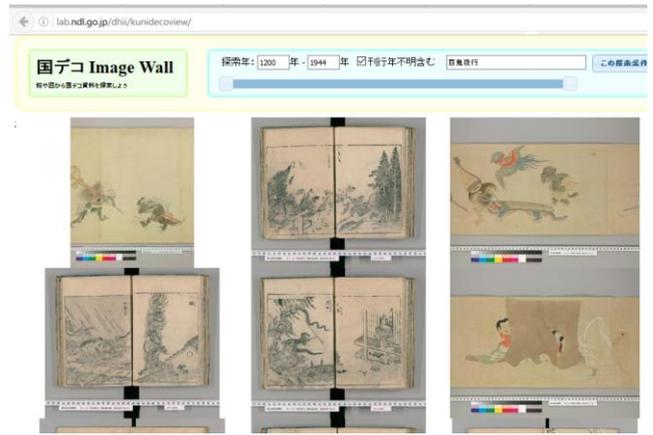
絵図一覧用のサムネイル画像は、国デコ Image Wall サーバ上に Pyramid Tiff 画像で置かれている画像に対して、IIP Image server によって IIIF Image API 形式で縦横最大 300 ピクセルのサムネイル画像を生成・取得するようになっている。(図 1)



(図 1 国デコ Image Wall の絵図画像表示画面)

一覧される絵図画像は、年号とキーワードで絞り込むことができる。年号は、jQuery UI のスライダウィジェットによる指定と、数値を直接入力することによる指定という二つの指定方法が提供されている。キーワードは、NDL サーチにおいて提供されている書誌情報を図書資料のタイトル毎に取得したデータを用意しており、キーワードによる絞り込みが行われた場合、このデータを検索して該当する図書資料を選定し、その図書資料に含まれる絵図画像を表示するようになっている。たとえば、図 2 は、「百鬼夜行」で絞り込んだ絵図画像一覧の画面の一部だが、国デコ Image Wall に収録されている三種類の百鬼夜行関連古典籍資料の絵図画像がリストされている。

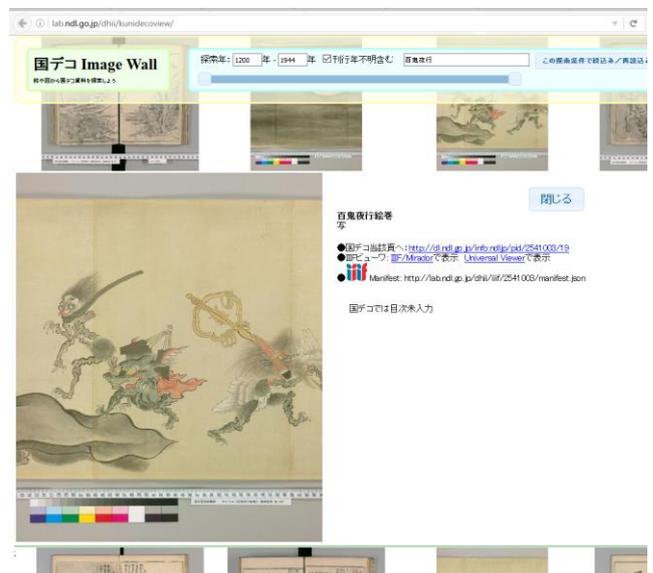
d なお、本講執筆時点では、国デコ中の古典籍には DOI が付与されており、「DOI/画像番号」という記法も可能である。



(図 2 「百鬼夜行」での絞り込み表示の一部)

このようにして表示させた画像の中から、気になるものをクリックすると、その画像が拡大表示され、メタデータ等も同時に表示され、図書資料へのリンクが提示される。ここでのリンクには 4 種類が用意される。(図 3)

1. 国デコの提供する図書資料ビューワにおける当該ページの表示画面
2. IIIF 対応ビューワ Mirador における当該ページの表示画面
3. IIIF 対応ビューワ Universal viewer における当該ページの表示画面
4. 当該資料・当該ページの IIIF Manifest のリンクを含む IIIF Drag & Drop アイコン



(図 3 任意の画像をクリックして拡大した状態)

通常は 1. をクリックすることで国デコの標準的な閲覧画面で表示して図書全体を見ていくことができるようになっている。それに加えて、国デコ Image Wall では、掲載している資料を IIIF に対応できるように変換・格納しているた

e <http://www.appelsiini.net/projects/lazyload>

め、2~4のように、IIIF 対応での表示を試行できるようになっている。4. については、YouTube にて実際の使い方を動画で提示しているため参照されたい(f. また、IIIF Manifest を提供しているため、他の IIIF 対応ビューワ上で表示することももちろん可能である。

また、いわゆる古典籍画像ではまだ確認できていないが、旧近代デジタルライブラリー由来の図書資料の中には章タイトルを入力しているものが相当数存在する。そのような資料に関しては、この任意画像の拡大表示画面では、章を表示して該当する章を検知して赤字とすることで、その図書のなかでその画像がどのような位置づけとなっているかを多少なりとも確認できるようにしている。(図4)



(図4 章タイトルを含む拡大表示画面)

以上が、国デコ Image Wall の概要である。原稿執筆時点では、今のところは大部分が古典籍だが、13世紀から20世紀前半の図書資料1,823冊で20,352件の画像が登録され、絵図から図書へと辿っていけるようになっている。では次に、図書の画像を抽出するための試みについて紹介したい。

### 3. Crowd4U による絵図画像判定

Crowd4U は、学術の文脈において、マイクロタスクによるクラウドソーシングを実現することを目指したプロジェクトであり、すでに様々なプロジェクトが展開されている。ここでは、Transcribe JP との連携の一環として、絵図画像かどうかの判定をするマイクロタスクを用意し、その結果としての画像 URL を国デコ Image Wall に提供している。現在、このタスクは、Web 上(g, Android スクリーンロックアプリケーション(h)に加えて、人が廊下を通った時に通る場所で判定するという等の仕組み(図5)で実現している。この詳細は[2]で報告しているが、結果として、多くの人がそれほど手間をかけずに参加して(g 絵図

### 画像 URL を収集



(図5 絵図画像判定タスクの例)

できるようになっている。今後もこれを継続していくことで、収録絵図画像件数の増加につなげていきたい。

### 4. 課題と展望

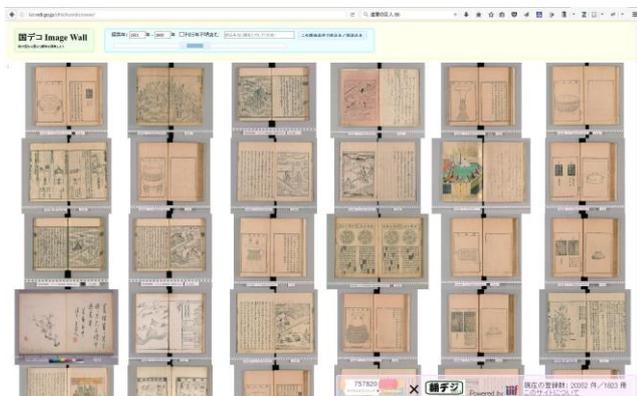
国デコ Image Wall を構築したことにより、すでに多くの人が認識していたであろう問題のいくつかが改めて浮き彫りとなった。まず、年号による絞込みは、国デコ資料におけるメタデータの付け方の問題に直面した。国デコ資料では、100年以上の期間で年号が不明な場合などに、年号のデータを1000としている。したがって、年号で検索できるシステムを用意した際、たとえば、1000年で検索すると美しい色刷りの浮世絵が検索結果として表示されるということになった。当面の回避策として、1000年を「刊行年不明」として位置づけ、それを検索結果に含むかどうかをチェックボックスで選択できるようにすることで、とりあえず多少の状況改善を試みた。しかしながら、100年単位で同定可能だがそれ以上詳しい年号がわからない場合には1200年、1300年、といった年号が用いられており、100年単位のものについてはそのままにしまっているため、インターフェイスとしてもう少し改善が必要になると思われる。また、データとしても、W3CDTF形式で1000となっているものでも issued のところに様々な形式で何らかの記述がなされていることが少なくない。江戸時代、江戸中期、江戸後期等、ある程度の幅のある期間指定となっているため、これは、たとえば、TEI Guidelines で規定されている notBefore, notAfter などといった属性を用いた記法を導入することである程度まで解決可能であると思われる(i)。キーワードによる絞込み機能も提供しているが、国デコのメタデータにおける漢字の正規化がそれほど行われておらず、たとえば出版年の項目でも、書写を表すと思われる

f <https://www.youtube.com/watch?v=zapU9hv-mo4>  
g <https://crowd4u.org/projects/crowd#hondigi>

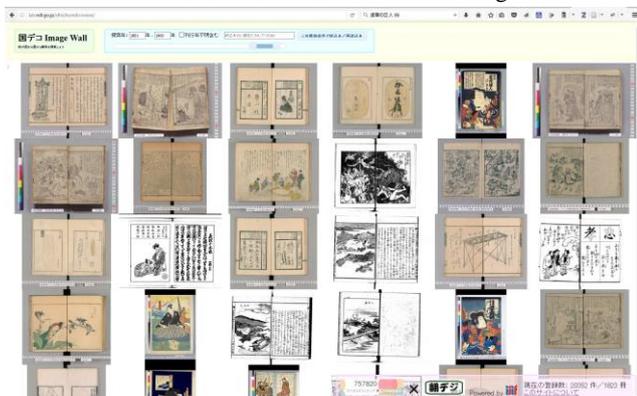
h <https://play.google.com/store/apps/details?id=org.crowd4u.lockscreen>  
i <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ND.html>

文字として「写」と「寫」が併用されているなど、漢字の異体字検索の仕組みを導入することが必要であると思われる。これは、今後導入していきたい。

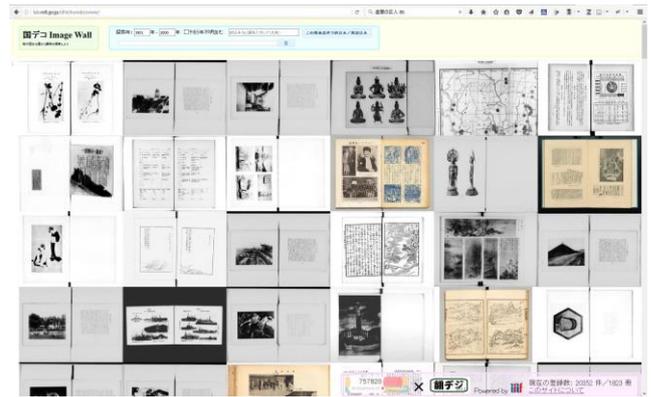
この仕組みは、あくまでも、絵図画像との偶然に近い出会いを通じて図書資料を探索するというものであり、人文学の研究に直接有益なものを提供する形にはなっていない。しかしながら、むしろ、研究者ではない人、あるいは、直接それを専門としない研究者が、必ずしも十分な前提知識を持たないままで探索して絵図画像に出会い、そこからまとまりのあるコンテンツへと辿っていくことを可能とする仕組みであり、工夫次第では、より深いコンテンツへと導いていき、そこから新しいものを生み出していくための入り口ともなり得るものである。このように考えるなら、オープンサイエンスにおいて期待されるシティズンサイエンスを実現するにあたっては、むしろ基盤の一つともなり得るものである。そして、IIIF 対応デジタル画像リポジトリの中にはアノテーションや翻刻を提供しているものもあり [3]、そういったリポジトリへもたどっていきけるようにすることで、さらに知識を広げ、人文学への関心を深めてもらうきっかけとすることも可能だろう。また、国デコ資料のコレクションとしての偏りから現状では十分とは言えないものの、ここで提供されている年号での絞込みは、時代毎の図書中の絵や図表に関する技法や流行の変遷についての理解に貢献することも可能だろう。その例として、16 世紀、19 世紀、20 世紀の画面を挙げておきたい。(図 6, 7, 8)



(図 6 16 世紀で絞り込んだ国デコ Image Wall)



(図 7 19 世紀で絞り込んだ国デコ Image Wall)



(図 8 20 世紀で絞り込んだ国デコ Image Wall)

そして、この仕組みは、通常のデジタルアーカイブ検索システムに組込むことも十分に可能であり、それによって検索システムの導線を増やすこともできるだろう。そのようにしてデジタルアーカイブへの導線を増やすことは、多様な利活用の可能性を高めていくことにつながり得る。今後は、上述のように、収録画像・図書資料を増やしていくとともに、この仕組みによる実際の効果についても調査を進めていきたい。

#### 謝辞

国デコ Image Wall は、国立国会図書館、特に次世代システム開発室の皆様のご支援・ご協力によって構築運営されていることを感謝とともに記しておきたい。また、国デコ Image Wall における古典籍の画像 URL 収集にあたっては、中村美里氏のご協力をいただいたので感謝して記しておきたい。なお、本発表の一部は、国立情報学研究所公募型共同研究「文化資料デジタルアーカイブの研究活用を志向するフレームワークの研究」の助成、および JSPS 科研費 (JP25240012, JP15H05725) を受けて遂行されたものである。また、Crowd4U に貢献されている全ての方に感謝申し上げる。一部の方のお名前は crowd4u.org にあるが、特に、筑波大学附属図書館には多大な貢献をいただいている。また、匿名の多くの貢献者の皆様にも感謝申し上げます。

#### 参考文献

- [1] 岡本真・柳与志夫責任編集『デジタル・アーカイブとは何か』勉誠出版, 2015.
- [2] 太田千尋, 森嶋厚行, 寺田努, 中村聡史, 北川博之, “床に投影したマイクロタスクのデータ品質向上手法”. 第 8 回データ工学と情報マネジメントに関するフォーラム (DEIM2016), (2016/2/29-3/2), 7 pages.
- [3] Kiyonori Nagasaki, Tetsuei Tsuda, Charles Muller, Masahiro Shimoda, “Tagging on Buddhist Images via IIIF and TEI encoding”, Text Encoding Initiative Conference 2016, Vienna (Austria), (2016/9/26-28), pp. 141-143.