# 資料画像公開・利用の国際化と高度化の取り組み：「日本語史研究資料［国立国語研究所蔵］」の事例

国立国語研究所論集 (NINJAL Research Papers) 15: 163–176 (2018)
ISSN: 2186-134X print/2186-1358 online


# 資料画像公開・利用の国際化と高度化の取り組み

## ――「日本語史研究資料［国立国語研究所蔵］」の事例――

高田智和[a]　福山雅深[b]
堤　智昭[c]　小助川貞次[d]

[a] 国立国語研究所 研究系 言語変化研究領域
[b] 東京農工大学 博士前期課程
[c] 東京電機大学／国立国語研究所 共同研究員
[d] 富山大学／国立国語研究所 共同研究員

**要旨**

昨今，デジタル技術の進歩，学術政策におけるオープンサイエンス・オープンデータの推進と相まって，世界各国の様々な機関が所蔵する資料（主に古典籍）のデジタル画像化とインターネットを通じた公開が進んでいる。従来の公開では，単純な閲覧を目的とした場合，JPEG 形式や TIFF 形式のデジタル画像を提供する方式や専用ビューアを提供する方式が採用されてきた。また，アノテーションが付与された画像を表示する場合や，複数画像を比較表示するような場合，提供画像と専用ビューアを合わせて用意する方式が採用されてきた。本稿では，従来の公開方式による「日本語史研究資料［国立国語研究所蔵］」に，IIIF 規格に準拠した画像公開システムを導入した事例と，コーパス検索結果からの画像参照への実装を報告する *。

**キーワード：**古典籍，デジタルアーカイブ，資料公開，IIIF，コーパス

### 1. はじめに

21 世紀に入ってから，国内外の図書館，博物館，大学，研究機関において，所蔵資料（主に古典籍）のデジタル画像公開が進んでいる。一般に「デジタルアーカイブ」と呼称される Web コンテンツである。デジタル技術の進歩と，オープンサイエンス・オープンデータの推進と相まって，「デジタルアーカイブ」は今後も拡張していくことであろう。

国立国語研究所でも 2012 年から，所蔵資料のデジタル画像化と公開を進めている（「日本語史研究資料［国立国語研究所蔵］」http://dglb01.ninjal.ac.jp/ninjaldl/，以下「国語研日本語史研究資料」）。2018 年 1 月現在 58 点を公開している。本稿では，（1）「国語研日本語史研究資料」の公開コンテンツについて紹介し，（2）公開画像の利用例としてコーパス検索結果からの画像参照の仕組みを解説し，（3）近年欧米の機関で主流となっている IIIF（International Image Interoperability Framework, http://iiif.io/）の「国語研日本語史研究資料」に対する導入，（4）全文検索システム「ひまわり」からの画像参照への応用について報告する。

* 本研究は，人間文化研究機構広領域連携型基幹研究プロジェクト「異分野融合による「総合書物学」の構築」の国語研ユニット「表記情報と書誌形態情報を加えた日本語歴史コーパスの精緻化」（プロジェクトリーダー：高田智和），科学研究費基盤研究（B）「字体記述のデジタル化に基づく文字規範史の定位」（課題番号 26284066，研究代表者：高田智和）による成果の一部である。

## 2.「日本語史研究資料［国立国語研究所蔵］」

「国語研日本語史研究資料」は，国立国語研究所研究図書室所蔵資料の中から，『国語学史資料集』（国語学会編，1979 年，武蔵野書院）に取り上げられたものや，「日本語歴史コーパス」の構築で利用できるものを優先的に選定し，デジタル画像と翻字本文テキストとを公開するものである。現在の公開資料 58 点（うち，テキストも公開している 7 点は書名の前に * を付与）は以下の通りである。

1. 金剛頂一切如来真実摂大乗現証大教王経（1 巻，院政期写本）
2. 金剛頂一切如来真実摂大乗現証大教王経（3 巻，嘉応二年加点本，院政期写）
3. 悉曇蔵（2 冊，1297[永仁 5] 年写）
4. 古今文字讃（3 巻，1503[文亀 6] 年写）
5. 古文真宝抄（12 冊，1525[大永 5] 年以降写）
6. 標題徐状元補注蒙求（3 冊，室町末期－江戸初期写）
7. 易林本節用集（1 冊，1597[慶長 2] 年以降刊）
8. 倭玉篇（慶長十八年版）（3 冊，1613[慶長 18] 年刊）
9. 尚書（古活字版）（3 冊，1596[慶長元] － 1615[慶長 20] 年刊）
10. 大学抄（1 冊，1615[元和元] 年写）
11. 絶句抄（3 冊，1623[元和 9] 年以前写）
12. 中華若木詩抄（寛永十年版）（1 冊，1633[寛永 10] 年刊）
13. かたこと（5 冊，1650[慶安 3] 年刊）
14. 徒然草（寛文七年版）（2 冊，1667[寛文 7] 年刊）
15. 増補下学集（5 冊，1669[寛文 9] 年刊）
16. 仮名文字遣（1 冊，江戸前期刊か）
17. しちすつ仮名文字使蜆縮涼鼓集（2 冊，1695[元禄 8] 年刊）
18. 補忘記（元禄版）（1 冊，1695[元禄 8] 年刊）
19. 唐話纂要（6 冊，1718[享保 3] 年刊）
20. 和字正濫鈔（5 冊，1739[元文 4] 年刊）
21. 磨光韻鏡（2 冊，1744[延享元 ] 年刊）
22. 聖遊郭（雪月花）（1 冊，1757[宝暦 7] 年刊）
23. かさし抄（3 冊，1767[明和 4] 年序）
24. *諸国方言物類称呼（5 冊，1775[安永 4] 年刊）
25. あゆひ抄（2 冊，1778[安永 7] 年刊）
26. 通言総籬（1 冊，1787[天明 7] 年序）
27. 大礒風俗仕懸文庫（1 冊，1791[寛政 3] 年刊）
28. 傾城買二筋道（1 冊，1798[寛政 10] 年序）
29. 青楼阿蘭陀鏡（5 冊，1798[寛政 10] 年刊）

30. 石場妓談辰巳婦言（1 冊，1798[寛政 10] 年序）
31. 蛮語箋（1 冊，1798[寛政 10] 年序）
32. 字音仮字用格（1 冊，1799[寛政 11] 年刊）
33. 訳鍵（1 冊，1810[文化 7] 年刊か）
34. 再考増補標註古言梯（1 冊，1820[文政 3] 年刊）
35. 河東方言箱枕（3 冊，1822[文政 5] 年刊）
36. 玉菊全伝花街鑑（3 冊，1822[文政 5] 年序）
37. *小三金五郎仮名文章娘節用（3 冊，1831[天保 2] − 1834[天保 5] 年刊）
38. *春色梅児与美（4 冊，1832[天保 3] − 1833[天保 4] 年刊）
39. *梅暦余興春色辰巳園（4 冊，1833[天保 4] − 1835[天保 6] 年序・刊）
40. *比翼連理花廼志満台（4 冊，1836[天保 7] − 1838[天保 9] 年序・刊）
41. 潮来婦誌（3 冊，江戸後期刊）
42. 仮字本末（4 冊，1850[嘉永 3] 年刊）
43. 仮字類纂（1 冊，1854[嘉永 7] 年刊）
44. 三語便覧（3 冊，1854[嘉永 7] 年序）
45. 和蘭字彙（10 冊，1855[安政 2] 年刊）
46. ゑんぎりしことば（1 冊，1860[万延元 ] 年序）
47. 音韻仮字用例（3 冊，1860[万延元 ] 年刊）
48. 英語箋（2 冊，1861[万延 2] 年刊）
49. *おくみ惣次郎春色江戸紫（7 冊，1864[元治元] −明治刊）
50. 牛店雑談安愚楽鍋（3 冊，1871[明治 4] − 1872[明治 5] 年序・刊）
51. 明六雑誌（43 冊，第 1 号 1874[明治 7] 年 3 月 − 第 43 号 1875[明治 8] 年 11 月刊）
52. *哲学字彙（1 冊，1881[明治 14] 年刊）
53. 東洋学芸雑誌（87 冊，第 1 号 1881[明治 14] 年 10 月 − 第 87 号 1888[明治 21] 年 12 月刊）
54. 改正増補和英英和語林集成（1 冊，1886[明治 19] 年刊）
55. 国民之友（70 冊，第 1 号 1887[明治 20] 年 2 月 − 第 70 号 1890[明治 23] 年 1 月刊）
56. 尋常小学読本（国定読本第 1 期）（8 冊，1904[明治 37] − 1906[明治 39] 年刊）
57. 音韻分布図（1 冊，1905[明治 38] 年刊）
58. 口語法分布図（1 冊，1906[明治 39] 年刊）

「国語研日本語史研究資料」はデジタルアーカイブとしては後発であり，コレクションの規模も極めて小さいものである。しかし，画像公開として一般的な見開き 1 丁（2 ページ）ではなく，片面半丁の 1 ページ単位での JPEG 画像と PDF 画像を提供している。後述するように，コーパス検索結果からの画像参照時に，ユーザーが検索語句を画像内から探す時間を多少なりとも短縮できるようにするためである。なお，JPEG 画像と PDF 画像は，CC BY-NC 4.0（クリエイティブ・コモンズ 表示 − 非営利 4.0 国際ライセンス）の下で提供している。画像利用の際には，資料

名と「人間文化研究機構国立国語研究所所蔵」であること表示すれば，学術利用での転載・加工にあたって利用申請は無用である。また，加点資料など細部にわたって拡大閲覧が必要な資料については，高精細画像ビューア ContentsView FLEX[1] によるサービスも行っている（図 1）。

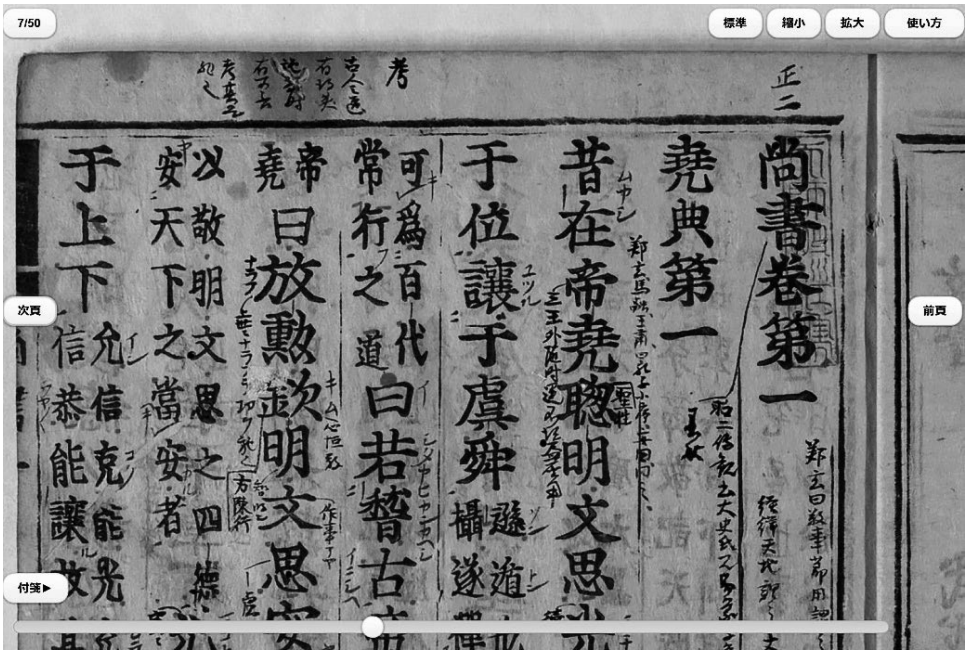


図 1　高精細画像ビューア ContentsView FLEX による表示例

## 3. コーパス検索結果からの画像参照とその課題

### 3.1 画像参照方式

「国語研日本語史研究資料」では，画像閲覧以外に他のシステムからの画像閲覧・取得を目的としたアンテナサイトを用意している。アンテナサイトは，全文検索システム「ひまわり」[2] やコーパス検索アプリケーション「中納言」[3] による検索結果から，画像参照を行う場合に利用する。ここでは，「ひまわり」からの利用例を挙げながら，アンテナサイトの構成について述べる。アンテナサイトは図 2 に示すように，画像を要求するシステムと原本画像を管理する資料情報 DB・原本画像ファイルサーバとの間を仲介するように機能する。アンテナサイトに，取得したい原本

[1] 岡山市・コンテンツ株式会社開発のビューア（http://www.contents-jp.com/demo/index.html）。「国語研日本語史研究資料」では「金剛頂一切如来真実摂大乗現証大教王経（院政期写本）」「金剛頂一切如来真実摂大乗現証大教王経（嘉応二年加点本）」「悉曇蔵」「古今文字讃」「尚書（古活字版）」の 5 点で利用している。また，国立国語研究所の公開コンテンツでは，「米国議会図書館蔵『源氏物語』画像（桐壺・須磨・柏木）」（http://dglb01.ninjal.ac.jp/lcgenji_image/）にも利用している。

[2] 最新版（安定版）は 2018 年 1 月 5 日公開の ver.1.5.7（http://www2.ninjal.ac.jp/lrc/index.php?%C1%B4%CA%B8%B8%A1%BA%F7%A5%B7%A5%B9%A5%C6%A5%E0%A1%D8%A4%D2%A4%DE%A4%EF%A4%EA%A1%D9）。

[3]『現代日本語書き言葉均衡コーパス』や『日本語歴史コーパス』を検索するための登録制 Web アプリケーション（https://chunagon.ninjal.ac.jp/auth/login）。

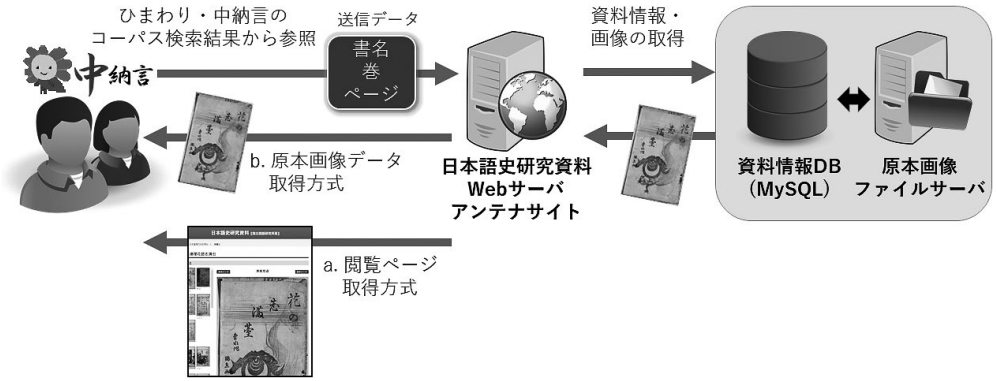


図 2　画像参照の概念図

画像の「書名」「巻」「ページ」をアクセス URL に付与する形で送信すると，アンテナサイト内で原本画像ファイルサーバから原本画像を取得するための処理が行われ，画像を要求したシステムにデータが提供される。またアンテナサイトは，閲覧ページの URL データを取得する方式と，原本画像データを取得する方式の 2 種類を用意している。それぞれの方式は以下のような用途を想定している。

・閲覧ページの URL データ取得方式
該当原本画像データを，インターネットブラウザを用いて表示する場合に使用する。他システムはインターネットブラウザで URL を開くためのリンクのみを提供するような用途を想定している。

・原本画像データ取得方式
該当原本画像データを，他システムが直接利用する場合に使用する。検索結果内での画像表示や画像の一括ダウンロードなどを想定している。

例えば,「ひまわり」を用いて検索した文章に対応する原本画像を「ひまわり」内で表示したり，原本画像を参照するためのリンクを「ひまわり」内に表示したりする利用が可能である。

本システムと連携している「人情本コーパス」[4] を，「ひまわり」を用いて利用した場合，「ひまわり」内で「国語研日本語史研究資料」内に保存されている原本画像の参照が可能である。「人情本コーパス」には，「国語研日本語史研究資料」からアンテナサイトを通じてデータを取得するために必要な「書名」「巻」「ページ」が XML タグの中に保存されている。「人情本コーパス」の場合，「書名」「巻」は XML の text タグに「<text title="比翼連理花迺志満台" volume="初編上" …. >」のように属性「title」に「書名」を，属性「volume」に「巻」を記述している。また，該当する書籍を閲覧するための URL を属性「url」に「url="http://dglb01.ninjal.ac.jp/ninjaldl/

[4]『日本語歴史コーパス 江戸時代編』の一部（http://pj.ninjal.ac.jp/corpus_center/chj/edo.html）。

show.php?title=hananosimadai"」のように記述し，アクセス時に利用している。「ページ」は，ページの区切りに pb タグを付与し，「<pb n="一オ" num="11">」のように属性「n」には丁数を，属性「num」には通しのページ番号をそれぞれ別に記述している。「ひまわり」は，これらの情報を用いて「http://dglb01.ninjal.ac.jp/ninjaldl/show.php?title=hananosimadai &issue =001 &num=11」のようにアクセス URL を生成する。URL では，パラメータ「title」に「書名」を，「issue」に「巻」を，「num」に「ページ」を指定する。このように「ひまわり」が，ユーザーの検索内容に応じて必要な情報をアンテナサイトに送信し，「ひまわり」内部で使用する場合には原本画像を取得し表示する。また，ユーザーが原本画像の詳細な情報を得たい場合を想定し，閲覧ページの URL データを表示した画像にリンクとして埋め込んでおり，画像をクリックすることで「国語研日本語史研究資料」の該当ページが表示可能である。

### 3.2 画像参照における課題

3.1 で述べた現行の画像参照方式では，取得した原本画像がページ単位で表示されるため，検索語が原本画像のどこに存在するのかが分かりにくいという問題がある。この問題の解決策として，画像内にアノテーションを付与する方法が考えられる。

しかし，従来のアノテーション付与方式では，どういったシステムでアノテーション付与するかによって固有の問題が発生すると考えられる。ここでは，①原本画像をダウンロードした他システム（「ひまわり」など）でアノテーションを付与する方式と，②原本画像を公開している「国語研日本語史研究資料」のサーバが独自の方法で画像加工を施しアノテーションを付与する方式の二つの方式において発生する問題について述べる。

①の方式では，原本画像を利用する他システムが自由にアノテーションを付与できるため，それぞれのシステムに合わせた柔軟な画像利用が可能となる。一方，他システムを開発・管理する側がアノテーションを付与するためのシステムを開発し，管理しなければならないため負担が増大し，結果として「国語研日本語史研究資料」の利用を阻害する要因となりえる。

②の方式では，他システムはアノテーションが付与済みの画像を利用することになるため，簡単に利用できるという利点がある。一方，他システムに適した柔軟な画像利用は難しく，画像を提供する側の開発・管理負担も増大する。また，アノテーションを付与した画像を閲覧するために独自のビューアなどを用いる場合，利用者が独自ビューア独特の操作性・機能を理解しなければならず，利用するためのハードルが高くなる。その結果として，「国語研日本語史研究資料」の利用を阻害する要因となりかねない。このように，従来の方式ではアノテーションが付与された画像の提供・管理コストが非常に大きいのが問題であると言える。

## 4. IIIF（International Image Interoperability Framework）について

### 4.1 IIIF の普及

3.2 で示した画像参照の課題を解決するため，「国語研日本語史研究資料」のシステムに IIIF を導入する。IIIF は，様々な機関が所蔵する画像コレクションをオープンデータとし，サーバ・

ビューア間で画像をやりとりする方式を共通化するために開発された規格である（永崎 2016）。これを用いることで，画像を提供する側の開発・管理負担を軽減し，かつ利用者が独自ビューア独特の操作性・機能を理解しなければならないといった負担を軽減したアノテーション付きの画像公開が可能となる。

IIIF の開発プロジェクトには世界各地の学術機関や図書館などが参画し，規格策定が進められている。国外においては，スタンフォード大学図書館，ハーバード大学美術館，大英図書館，フランス国立図書館などが，所蔵する書物や美術品等の公開に利用しており，今後のさらなる普及が期待される。一方，国内では徐々に関心が高まっているのが現状である（北本・山本 2016，北本 2017，佐藤・太田 2017）。国内での導入例として，人文情報学研究所の「国文研データセット簡易 Web 閲覧」（http://www2.dhii.jp/nijl_opendata/openimages.php），国文学研究資料館と人文学オープンデータ共同利用センターの「日本古典籍データセット」（http://codh.rois.ac.jp/pmjt/），東日本大震災アーカイブ福島協議会の「東日本大震災アーカイブ Fukushima」（http://fukushima.archive-disasters.jp/doc/），国文学研究資料館の「新日本古典籍総合データベース」（https://kotenseki.nijl.ac.jp/）などが挙げられる。

### 4.2 IIIF が解決する課題

現在インターネット上で公開されているデジタルアーカイブのほとんどは，それぞれ専用に独自開発されたビューアを使用することを前提としており，他のビューアを使ってアクセスすることはできない（図 3）。このことによって引き起こされる問題を二つの観点から例示する。

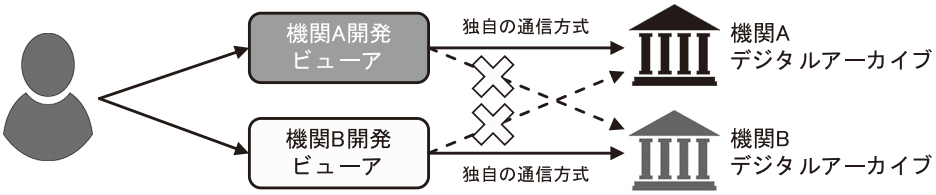


図 3　従来のビューアとデジタルアーカイブの関係

一つ目は，画像を閲覧する際の不便性である。多くの場合，デジタルアーカイブの閲覧に使用するビューアは公開機関等がそれぞれ独自に開発したものであるため，ビューアによって操作性が大きく異なり，機能にも差があるため，利用者は毎回操作を習得することが求められる。二つ目は，画像データを利活用する際の不便性である。例えば古典籍では，一つの書物に異本があり，それらが複数の機関に分散して所蔵されていることも多く，複数の異本を一つのビューア内で比較閲覧したいという研究者の需要もある。ところが，現在のデジタルアーカイブでは，サーバ・ビューア間の画像の通信方式が独自の仕様となっているため，一つのビューアで複数機関からの画像の取得に対応することは困難である。

これらの問題点を解決するために策定された規格が IIIF である。IIIF では，ビューアからサーバへ画像を要求するときと，サーバからビューアへ画像を提供するときの通信手順・通信内容を

定めている。これにより，デジタルアーカイブ提供機関のサーバが IIIF に対応すれば，任意の IIIF 対応ビューアから，あらゆるデジタルアーカイブにアクセスできるようになる（図 4）。また，書名，著者名などの関連情報を画像と紐づけてやり取りする仕組みも定められている。

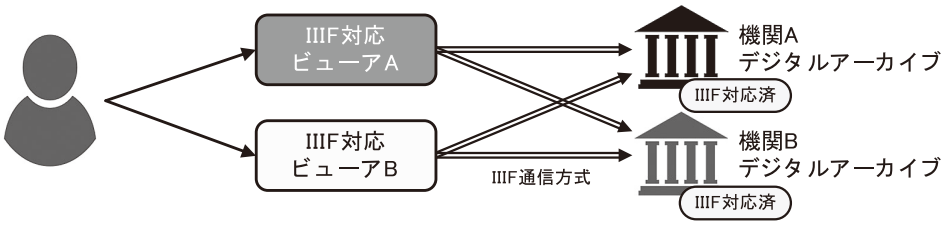


図 4　IIIF ビューアとデジタルアーカイブの関係

### 4.3 IIIF による画像提供

IIIF 規格でデジタルアーカイブを提供するには，IIIF に対応した画像提供サーバを設置し，マニフェストと呼ばれるファイルとともに公開する必要がある。マニフェストは，書名，著者名，ライセンス，各ページの画像データ URL やページ名などの情報が含まれている。そのため，多くの場合書物 1 冊に対して 1 ファイルのマニフェストを作成する。画像利用者はデジタルアーカイブを提供する機関のウェブページなどを通じてマニフェストを取得し，任意の IIIF 対応ビューアに読み込ませることで，デジタルアーカイブを閲覧・操作することが可能となる。また，公開機関があらかじめビューアに特定のマニフェストを読み込ませた状態で，利用者にビューアを提供することも可能である。さらに，アノテーションファイルを作成し，マニフェストファイルと紐づけることで，ページと位置を指定して画像にアノテーション情報を付与することも可能である。一部の IIIF 対応ビューアはマニフェストからアノテーション情報を読み込み，該当位置に表示する機能に対応している。

### 4.4 IIIF による画像閲覧

4.3 でも述べた通り，IIIF 形式で提供されるデジタルアーカイブは，IIIF 対応のビューアを用いることで閲覧が可能となる。IIIF 対応ビューアは世界中で開発が進められており，Mirador，OpenSeadragon，Leaflet-IIIF，IIIF Curation Viewer などがある。その多くはオープンソースとなっており，有志によって改善の提案や不具合の修正が頻繁に行われている。ビューアによっては画像を比較表示する機能や画像内の位置情報取得機能を持つものなど，機能や使い勝手が異なるが，いずれも IIIF 規格に準拠しているため，利用者が用途や好みに応じてビューアを選択し，デジタルアーカイブにアクセスできることが大きな特徴である。

## 5. 国立国語研究所における IIIF の導入

### 5.1 システム構成と画像閲覧方法

「国語研日本語史研究資料」では，2016 年 12 月から IIIF を導入し，画像の提供，マニフェストの公開，アンテナサイトの改良，IIIF 対応ビューアの改良を行った。4 で述べた通り，IIIF 規格は国内外で利用が進んできているデジタルアーカイブを提供するためのフレームワークである。画像公開サイトへ IIIF を導入することで，国際的に共通の規格で資料画像を公開し，システムの国際化と高度化を図った。

IIIF の導入にあたり，次に示す 4 点を「国語研日本語史研究資料」サーバに設置した。

（1）ピラミッドタイル TIFF 形式の原本画像ファイル
（2）IIIF 対応画像配信サーバプログラム「IIPImageServer」
（3）IIIF マニフェストファイル（アノテーションデータを含む）
（4）IIIF 対応ビューア「Mirador」

これにより，図 5 に示すような構成で，IIIF に準拠した画像公開が可能となる。具体的には，IIIF に対応した画像ビューアが，「国語研日本語史研究資料」が公開しているコンテンツのマニフェストファイルを読み込み，その内容に従ってサーバから画像データとアノテーションに関するデータを取得・表示する。

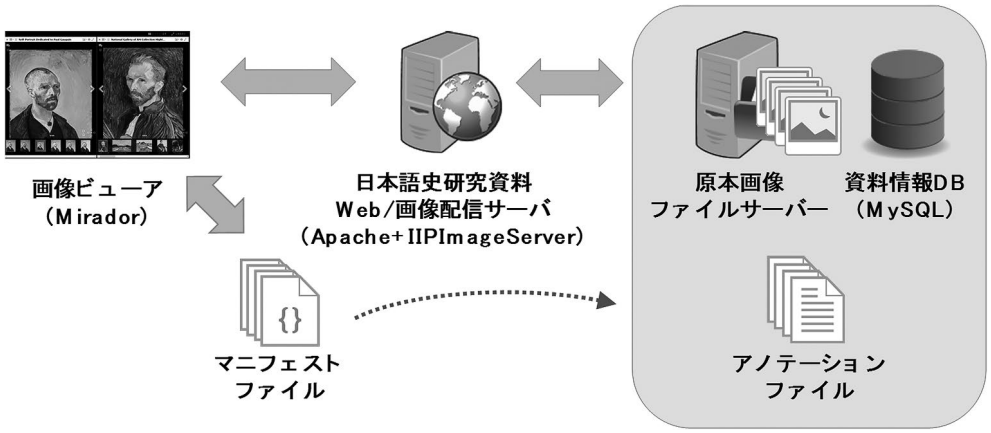


図 5 「国語研日本語史研究資料」のシステム構成概念図

（1）ピラミッドタイル TIFF 形式の原本画像ファイル

ピラミッドタイル TIFF 形式は，IIIF で画像を提供する時に用いる画像保存形式である。この形式では，一つの実体ファイルの中に解像度の高い画像から低い画像まで，解像度の異なる複数の画像が含まれる。IIIF 対応のビューアは拡大縮小に対応しており，この形式のデータを使うことで画像をダウンロードして表示するまでにかかる時間とシステムの利用効率を改善できる。縮

小して表示している時には解像度の小さいファイルを取得し，拡大して表示している時には解像度の高い画像を取得することで，特に高い解像度でファイルサイズの大きい画像を閲覧する時に，システムのスムーズな動作が可能となる。

（2）IIIF 対応画像配信サーバプログラム「IIPImageServer」

IIPImageServer は軽量で高機能な画像配信サーバプログラムであり，Apache や Lighttpd などの一般的な Web サーバ上で使用が可能である。このプログラムは，サーバ上にあるピラミッドタイル画像をクライアント（ビューア）でのズーム率や表示箇所にもとづいてリアルタイムに変換し，IIIF の画像配信方式でクライアントに配信する。

（3）IIIF マニフェストファイル（アノテーションデータを含む）

IIIF マニフェストファイルには，資料のタイトルや写刊年，その資料に含まれる全ページの画像 URI などのメタデータが JSON 形式で格納されている。一つのマニフェストファイルが物理的な紙媒体の資料 1 冊に対応する。また，アノテーションファイルもこのマニフェストファイルから参照できるように記述されている。ユーザーは，後述する IIIF ビューアに資料のマニフェストファイルを読み込ませると，画像やそれに関連する情報を閲覧することができる。

（4）IIIF 対応ビューア「Mirador」

「国語研日本語史研究資料」では，IIIF 対応画像ビューアとして Mirador を提供する。Mirador はアノテーション表示機能を持っており，本サイトが提供するアノテーション情報にもとづいて閲覧画像にアノテーションを表示できる。後述する「ひまわり」からの画像参照向けである。

IIIF を導入した「国語研日本語史研究資料」では，利用者が任意の IIIF 対応ビューアを用いて，「国語研日本語史研究資料」の画像を読み込むことができる（国立国語研究所が提供する Mirador を使わなくても画像を閲覧することができる）。例えば，人文学オープンデータ共同利用センターが提供するビューア（IIIF Curation Viewer）にマニュフェストファイルを読み込ませれば，そのビューア上で「国語研日本語史研究資料」の画像を閲覧することができる。

### 5.2 IIIF を用いた他サービスとの連携—コーパス検索結果からの画像参照の改良—

IIIF を活用した他システムの連携として，「ひまわり」からの画像閲覧の高度化を図った。具体的には，図 6 に示すように Mirador のアノテーション機能を用い，「ひまわり」から「人情本コーパス」を検索し，画像を参照するときに検索語句が原本画像のどこに存在するのかを示した。アノテーションは画像上に該当行を中心とした矩形の枠として表示し，利用者は検索語句を含む行と前後 1 行の計 3 行が画像上のどこにあるのかを容易に見つけられるようにした。3 行とすることで，ページ全体よりも利用者の目視範囲が狭まる。なお，前後 1 行としたのは，検索語句が行をまたぐことがあるためである。

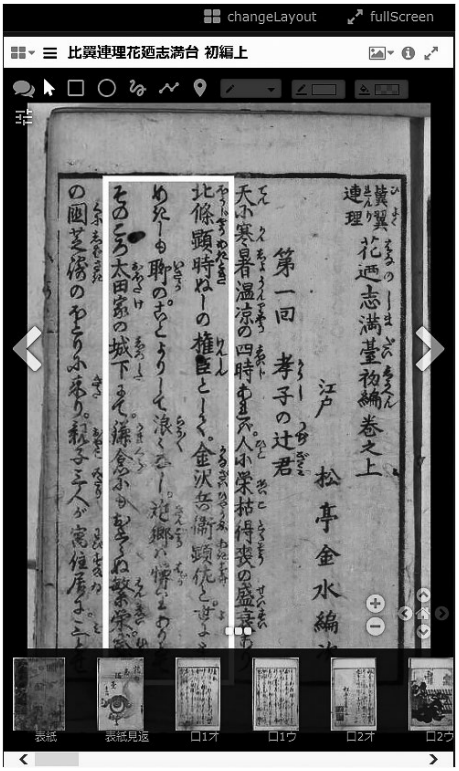


図 6　IIIF ビューアを用いた画像参照の例

「ひまわり」のコーパス検索結果など，他システムから画像をアノテーション付きで参照するときの処理手順を図 7（次頁）に示す。他システムは IIIF 用に改良したアンテナサイトを通じて，該当する画像データを閲覧できる状態になった Mirador へのアクセス URL を取得する。「ひまわり」の場合，まず「ひまわり」がアンテナサイトに「書名」「巻」「ページ」「行」のデータを URL に記述し送信する。具体的には，「http://dglb01.ninjal.ac.jp/mirador/annotation/hananosimadai/001/11/06」のような URL を生成する。この URL では，「http://dglb01.ninjal.ac.jp/mirador/annotation/」に続けて，パラメータとして「書名」「巻」「ページ」「行」の順に記述する。生成した URL を用いてアンテナサイトにアクセスすると，アンテナサイトが資料情報 DB と原本画像ファイルサーバから必要なデータを取得する。その後，アンテナサイトが「国語研日本語史研究資料」の資料閲覧用に国立国語研究所が提供している Mirador へ当該ページの画像と当該行のアノテーション情報を送信し，画像が閲覧できる状態になった Mirador へのアクセス URL を「ひまわり」へ送信する。「ひまわり」は，受け取った URL をインターネットブラウザで開くことで，アノテーション付きの画像を参照できる。

上記の処理を実現するためには，画像提供サーバの IIIF 対応に加え，アノテーションを付ける位置を記録したアノテーションファイルの作成と「人情本コーパス」への「行」情報の追記が必要となる。

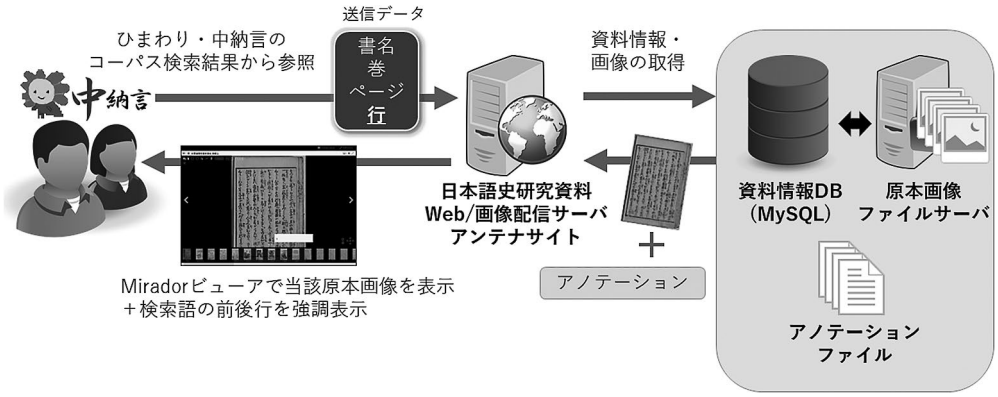


図 7　IIIF 導入後の画像参照の概念図

(1) アノテーションファイルの作成

本システムでは, マニフェストファイルとともにアノテーションファイルを作成する。アノテーションファイルには，書物の各行について，前後 1 行を含む計 3 行の位置情報を書き込む。この処理はアノテーション付与専用の IIIF ビューアを用いて，各巻の全ての画像について行う。この処理は，マウスを用いて人手で矩形の右上と左下，もしくは左上と右下を指定する。この処理については，公開する資料の数が多くなるほど多くの人手を要することになるため，自動化の検討が必要である。

(2)「人情本コーパス」への「行」情報の追記

「人情本コーパス」を修正し，検索語がページ内の何行目にあるかを示す「行」も出力するよう変更を行った。その結果,「ひまわり」で検索した場合，図 8 に示すように「書名」「巻」「ページ」などの従来の情報に加えて「行」も出力可能となった。この「行」の項目をダブルクリックすると「書名」「巻」「ページ」「行」が「国語研日本語史研究資料」サーバに送信され, インターネットブラウザが起動し Mirador で検索文字列を含む 3 行が矩形の枠で強調された状態で原本画像が表示される。

2017 年 11 月には，今回の新しい参照方式に対応したひまわり版「人情本コーパス」Ver.0.2 を公開した。これは国立国語研究所コーパス開発センターの Web ページからダウンロードすることができる（http://pj.ninjal.ac.jp/corpus_center/chj/edo.html#ninjou）。


図 8 「人情本コーパス」Ver.0.2 の「ひまわり」による検索結果

## 6. おわりに

「国語研日本語史研究資料」では，2016 年 12 月までに，それまでに公開していた資料のピラミッド TIFF 画像への変換と，マニフェストファイルの作成を完了し，IIIF を導入した。それ以降の画像公開は従来方式（JPEG 画像と PDF 画像）と IIIF 方式を併用することとし，現在 54 点が IIIF に対応している。今後の公開においても，従来方式（JPEG 画像と PDF 画像）と IIIF 方式の併用で画像提供を行う予定である。また，高精細画像ビューア ContentsView FLEX によるサービスも維持する。利用者の用途が様々であるため，公開方式にも多様性を持たせた方が良いと考えたからである。しかし，これには公開者側のコスト増が伴う。

本稿では，「国語研日本語史研究資料」の資料画像公開方式について解説し，利用の高度化を図るため IIIF 規格に対応した画像公開を行った事例について述べた。本事例では，IIIF を導入することにより，デジタルアーカイブと他システムとの高度連携が，従来よりも容易になった。管理・開発コストと利用者の利便性を両立できるようにするためには，共通のプラットフォームを利用することが重要である。IIIF は画像の共通プラットフォームとして有効であり，言語研究のシステムとの連携も可能である。歴史研究の場合，コーパスと原本画像との相互運用には高いニーズがあり，大局的には IIIF によって相互運用実現のコストダウンが図られたと言えよう。

## 参照文献

北本朝展（2017）「IIIF Curation Viewer の開発と利用」「IIIF Japan シンポジウム～デジタルアーカイブにおける画像公開の新潮流」（2017 年 10 月 17 日，於：一橋講堂）発表資料．http://agora.ex.nii.ac.jp/~kitamoto/research/publications/iiif17b-ppt.pdf

北本朝展・山本和明（2016）「人文学データのオープン化を開拓する超学際的データプラットフォームの構築」『人文科学とコンピュータシンポジウム論文集』117–124.

永崎研宣（2016）「国際的な画像の相互運用の枠組み IIIF について」政策会議「デジタルアーカイブの連携に関する実務者協議会（第 5 回）」（平成 28 年 10 月 31 日）配付資料．http://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_kyougikai/jitumu/dai5/gijisidai.html（2018 年 3 月 1 日確認）

佐藤正尚・太田一行（2017）「Mirador を利用したクラウドソーシングによるコラボレーション・システム―

人文学の共同研究における分析と理論の構築を支援するシステムの提案」『情報処理学会研究会研究報告』2017-CH-114: 1–6.

# Implementation of Image Disclosure System Using IIIF in NINJAL

TAKADA Tomokazu[a]　FUKUYAMA Masami[b]
TSUTSUMI Tomoaki[c]　KOSUKEGAWA Teiji[d]

[a]Language Change Division, Research Department, NINJAL
[b]Graduate Student, Tokyo University of Agriculture and Technology
[c]Tokyo Denki University / Project Collaborator, NINJAL
[d]University of Toyama / Project Collaborator, NINJAL

**Abstract**
Recently, progress in digital technology and the promotion of open science and open data in academic policies has led to a rapid digitalization of materials (mostly classical texts) from various institutions across the world and their publication on the Internet. Conventionally, digital data made public for the simple purpose of viewing have been presented in the JPEG or TIFF format or a specialized viewer. When displaying images with annotations or comparing several images, both the method of providing both images and the method of the specialized viewer are offered. This article reports on the case of "Collection of the NINJAL Research Library for Study of the Japanese Language History," an implementation of the image publication system based on IIIF specifications and the image references from corpus search results.

**Key words:** Japanese and Chinese classics, digital archive, disclosure system, IIIF, corpus