



第22回日本古典籍講習会（令和6年度）

国立国会図書館における 古典籍資料の電子化

利用者サービス部人文課

古典籍係 丹治美玲



目次

- 1.国立国会図書館デジタルアーカイブの変遷
- 2.国立国会図書館のデジタル化
- 3.古典籍資料のデジタル化
- 4.外注デジタル化の工程
- 5.内製デジタル化の工程
- 6.マイクロ資料からのデジタル化
- 7.まとめ

1. 国立国会図書館デジタルアーカイブの変遷



国立国会図書館が収集・保存しているデジタル資料の検索・閲覧

国立国会図書館のデジタルアーカイブの変遷とデジタル化の流れ

H8年度 電子図書館実証実験プロジェクト（館内限定）

H10.6 「デジタル貴重書展」公開

H12.3 「貴重書画像データベース」公開

H14.10 「近代デジタルライブラリー」公開

H18年度 「資料デジタル化基本計画」策定

H22年度 大規模デジタル化実施（H21年度補正予算）

H23.4 「国立国会図書館のデジタル化資料」公開

H23.10 大規模デジタル化撮影資料追加

H25年度 「資料デジタル化基本計画」策定

H26.1 「国立国会図書館デジタルコレクション」

R3.3 「資料デジタル化基本計画2021-2025」策定



デジタル貴重書展
H10（1998）公開
電子展示会

貴重書画像データベース
H12（2000）公開
検索機能つき





国立国会図書館のデジタル化資料

H23（2011）公開

公開当初は古典籍とインターネット資料のみで、図書は近代デジタルライブラリーから提供

その後、収録資料を増やし、図書、雑誌等とまとめて検索・閲覧可能になった（画面は2011年8月）



国立国会図書館デジタルコレクション

H26（2014）「国立国会図書館デジタル化資料」から名称変更
コンテンツの種類および収録資料数は年々増加

2. 国立国会図書館のデジタル化（1）

- ・ デジタル化の概要

国立国会図書館ビジョン2021-2025 –国立国会図書館のデジタルシフト–

https://www.ndl.go.jp/jp/aboutus/vision_ndl.html

資料デジタル化について

<https://www.ndl.go.jp/jp/preservation/digitization/index.html>

資料デジタル化基本計画2021-2025

https://www.ndl.go.jp/jp/preservation/digitization/digitization_plan2021.pdf

資料デジタル化実施計画2023-2025（内部計画）

デジタル化資料提供状況

<https://www.ndl.go.jp/jp/preservation/digitization/index.html>

図書：186万点

雑誌：137万点

2. 国立国会図書館のデジタル化 (2)

- ・マニュアル

国立国会図書館資料デジタル化の手引2017年版

<https://www.ndl.go.jp/jp/preservation/digitization/guide.html>

- ・遠隔研修

イントロダクション～資料デジタル化研修にあたって～

https://www.ndl.go.jp/jp/library/training/remote/digi_intro_2019.html

資料デジタル化の基礎

https://www.ndl.go.jp/jp/library/training/remote/digi_basic_2019.html

デジタル化資料の権利処理と利活用

https://www.ndl.go.jp/jp/library/training/remote/digi_copyright_2019.html

2. 国立国会図書館のデジタル化（3）

- ・ デジタル化の目的

- (1) 原資料の保存

- (2) 電子図書館サービスの実現

- (3) メタデータ充実等による資料の発見可能性向上

- (4) オープンデータ提供による新規コンテンツの創出

- (5) 災害による原資料散逸・破損への備え

参考：「国立国会図書館資料デジタル化の手引2017年版」
「資料デジタル化基本計画2021-2025」

3. 古典籍資料のデジタル化（1）

- ・ 古典籍資料室の所管資料

総計およそ29万冊

- ・ 江戸期以前の和古書
- ・ 清代以前の漢籍
- ・ 近代の新聞・雑誌のうち、準貴重書
- ・ 1830年以前刊行の西洋古典籍

デジタル化済み（約10万点）

- ・ 和古書・漢籍のうち貴重書約1,300点、準貴重書約800点
- ・ 準貴重書に指定された雑誌・新聞約130タイトル
- ・ 和古書・漢籍のうちコレクション資料など

3. 古典籍資料のデジタル化（2）

外注デジタル化

- ・ 受託業者が撮影・データ作成を行う
- ・ 撮影作業は原則館内、データ作成は受託業者施設
- ・ 年度ごとに調達・契約するため上半期調達、下半期撮影

内製デジタル化

- ・ 当館職員が撮影・データ作成を行う
- ・ 館内に設置されたスキャナを使用して撮影、ツールによりデータ作成
- ・ 通年で撮影

3. 古典籍資料のデジタル化 (3)

- ・資料デジタル化の優先順位

- (1) 貴重書・準貴重書等の和古書・漢籍
- (2) 貴重書・準貴重書以外の和古書
- (3) 貴重書・準貴重書以外の漢籍
- (4) 西洋古典籍 (番外)
- (5) 電子展示会で取り上げる資料 (番外)

- ・資料群別デジタル化方法

外注によるデジタル化 (1) (2) (3)

内製によるデジタル化 (2) (3) (4) (5)

4. 外注デジタル化の工程

1. 対象資料選定
2. 仕様書の作成
3. 業者決定（入札）
4. データ作成
（画像・メタデータ・管理データ）
5. 納品・検収
6. 画像とデータの登録・公開

4-1. 対象資料選定（1）

①対象資料の選定

ア．前年度に指定された貴重書・準貴重書

イ．それ以外の古典籍資料

※2020年度以前の規模は約30,000コマ/年

2021年度：70,000コマ

2022年度：75,000コマ

2023年度：80,000コマ

2024年度：80,000コマ



②現物の状態確認・コマ数確認

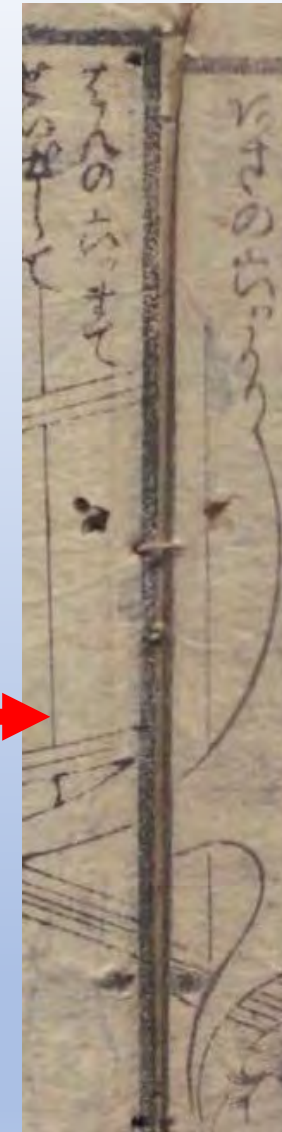


③対象資料決定・資料管理リスト確定

撮影前に綴じ糸を切る例
(改装本：帝国図書館表紙)



撮影前に綴じ糸を切る例 (改装本：帝国図書館表紙)



4-1. 対象資料選定 (2)

デジタル化対象資料のリスト作成

※H30年度の例

No.	枝番	請求記号	書誌ID	タイトル
12	1	わ025-12	024232504	群書一覽續篇

巻次	巻次よみ	各巻タイトル	各巻タイトルよみ	内容細目
巻三 有職之部	0003			

コレクション	大きさ	冊数	備考
	34cm	1冊	折込頁あり 間紙要(紙薄い) 付箋あり

メタデータ例

<https://dl.ndl.go.jp/pid/2569911/1/7>

The screenshot displays the National Diet Library Digital Collection interface. The main area shows a manuscript page with vertical Japanese text. To the right, a sidebar lists the contents of the volume, with '巻5-6' (Volume 5-6) selected. Below the manuscript image, the title '萬葉集 巻5-6' (Manyōshū Volume 5-6) is displayed. At the bottom, a metadata table provides detailed information about the document.

書誌情報	
資料種別	Book JapaneseClassicalBook
タイトル	萬葉集
タイトル読み	マンヨウシュウ
巻次又は部番号	巻5-6
巻次読み	5-6
並列タイトル	萬葉
並列タイトル読み	マンヨウ
出版年月日	【奈良年間】
出版年月日(W3COT)	1600

On the right side of the interface, there is a section for sharing and linking. It includes a '共有' (Share) button, a list of social media icons, and a '公開範囲' (Public Range) section. The '公開範囲' section indicates that the content is available for login-free viewing and is in the public domain (copyright expired). Below this, there is a '私蔵版の表示' (Display Private Version) section, which shows the specific URL for the digital collection item: <https://dl.ndl.go.jp/pid/2569911>, dated 2024-04-13.

巻号情報の例



巻号が番号のみで探しにくい
<https://dl.ndl.go.jp/pid/2610114>



帖の名前があり探しやすい
<https://dl.ndl.go.jp/pid/1288440>

電子化時の注記等の例 <https://dl.ndl.go.jp/pid/11891901>

職官考 (杉山叢書 ; 第33冊)

書誌情報 : 出版年月日 [1---] [写]

共有

この資料のURL
<https://dl.ndl.go.jp/pid/11891901>

公開範囲
ログインなしで閲覧可能
インターネット公開 (保護期間満了)
コンテンツの転載について

電子化時の注記
一部、料紙に重紙使用

DOI
[10.11501/11891901](https://doi.org/10.11501/11891901)

IIIF マニフェストURI
<https://dl.ndl.go.jp/manifest.json>

(参考) メタデータ流通ガイドライン
<https://iss.ndl.go.jp/information/guideline/>

4-2. 仕様書の作成 (1)

画像のフォーマット

- ・ 保存用画像

JPEG2000形式（可逆圧縮）

24ビットフルカラー

解像度は原資料に対して400dpi相当

※大型資料の全体画像は200～300dpi

- ・ 提供用画像（圧縮率1/50）

※「国立国会図書館資料デジタル化の手引|2017年版」 準拠

<https://www.ndl.go.jp/jp/preservation/digitization/guide.html>

4-2. 仕様書の作成（2）

撮影場所…国立国会図書館内

撮影作業員…経験者。特に指揮監督者は1級文書

情報管理士以上の資格を求める

撮影方法…平置き資料を垂直上方から撮影（原則）

資料の取り扱い…運び方、置き方、筆記用具等

写込み物…カラーチャート・メジャー・名票類

撮影順…分割・付箋・折込・挟み込み物等

※「国立国会図書館資料デジタル化の手引2017年版」に、

古典籍資料のものをもとにした仕様書例あり（pp.75-124）

写し込み物の例



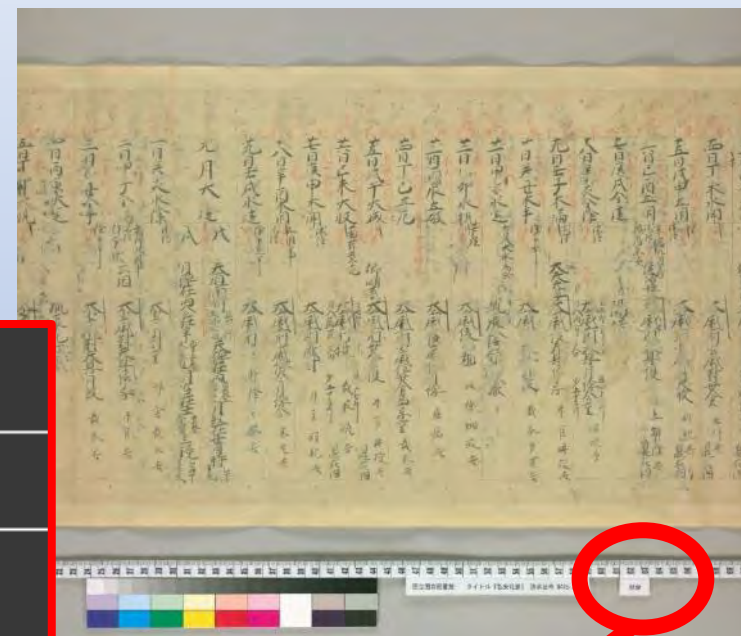
名票の例「紙背」 (本紙と紙背は各1分冊扱い)

<https://dl.ndl.go.jp/pid/2570133>

< 本紙 >



< 紙背 >



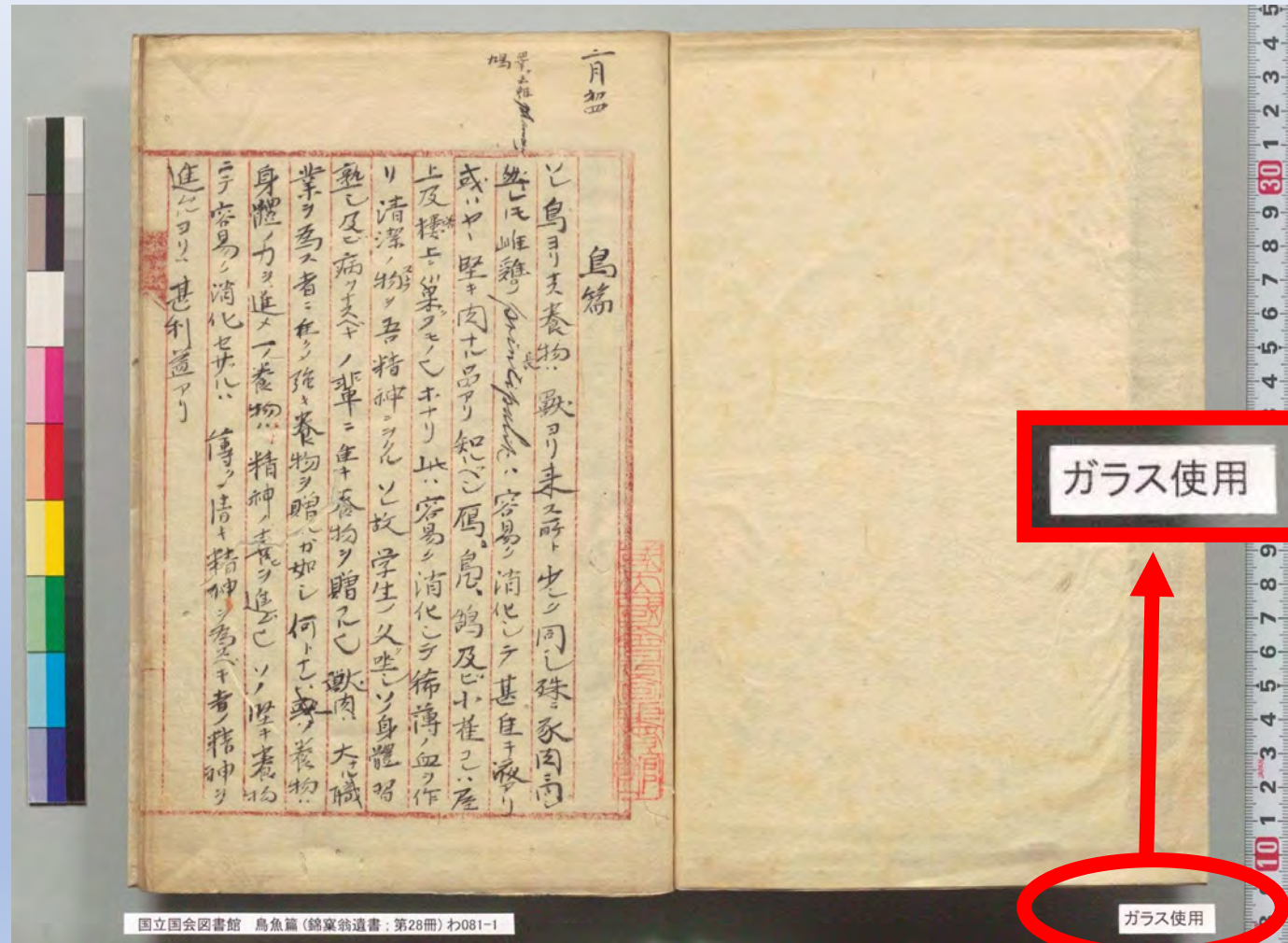
弘安禮節

> [1]

> [1]紙背

紙背

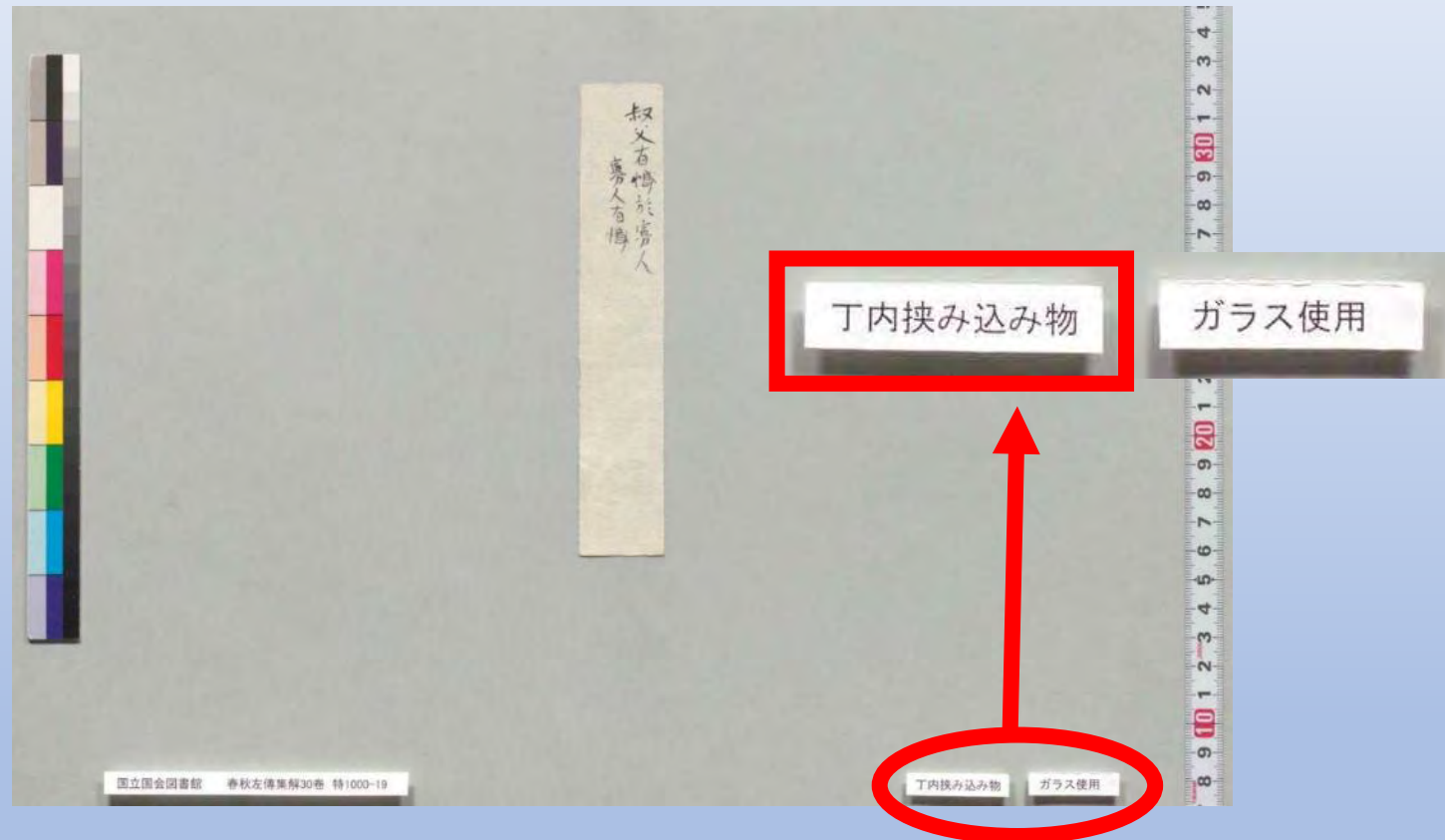
名票の例「ガラス使用」



ガラス不使用の例（貴重書・彩色がはがれやすいもの）

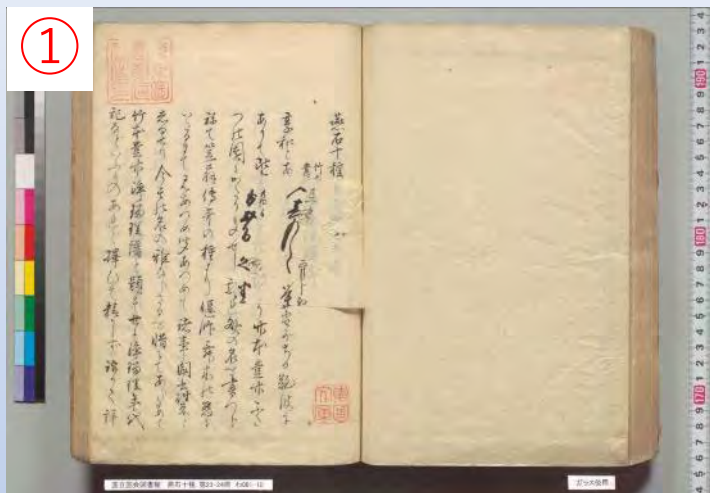


名票の例「丁内挟み込み物」



挟み込み物の例

<https://dl.ndl.go.jp/pid/11891859>



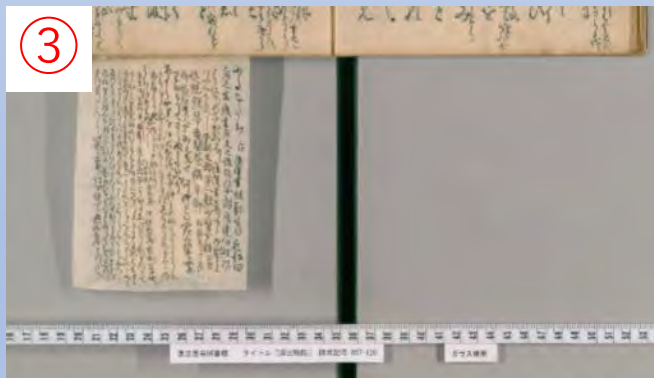
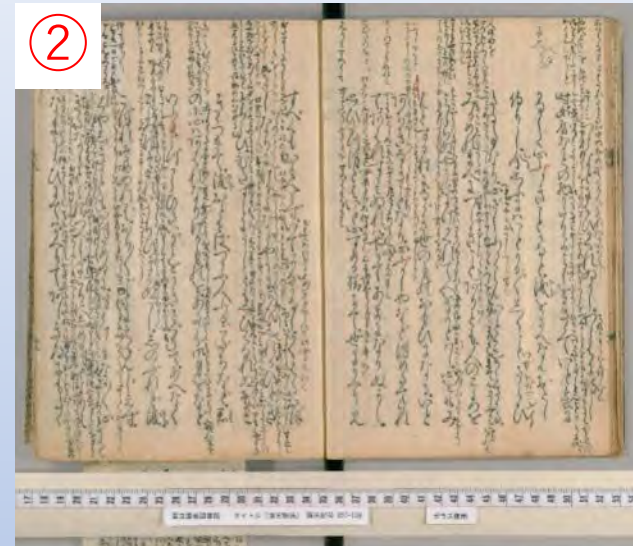
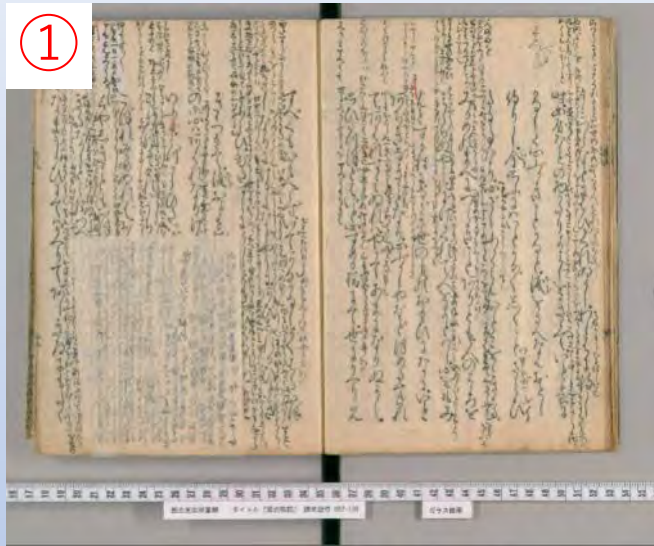
① 挟み込み物のある状態

② 挟み込み物のみ
「挟み込み物」の名票

③ 挟み込み物なしの状態

付箋の例

<https://dl.ndl.go.jp/pid/2579510/1/23>



① 付箋が折られた状態

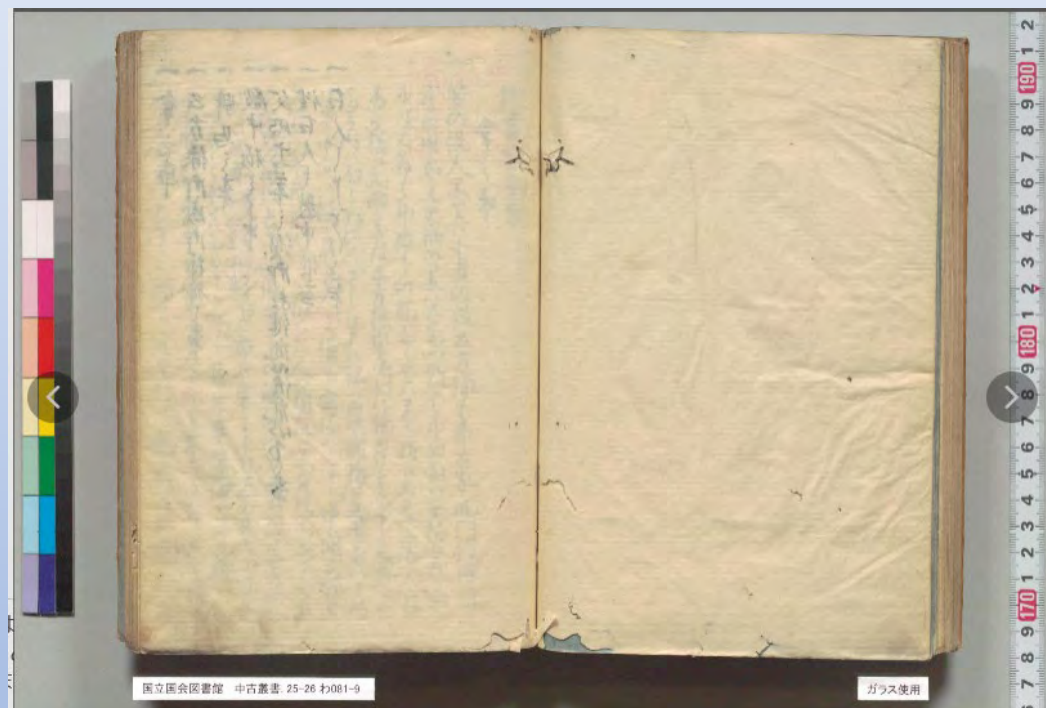
③ 付箋を下ろして資料全体

③ 下ろした付箋の全体

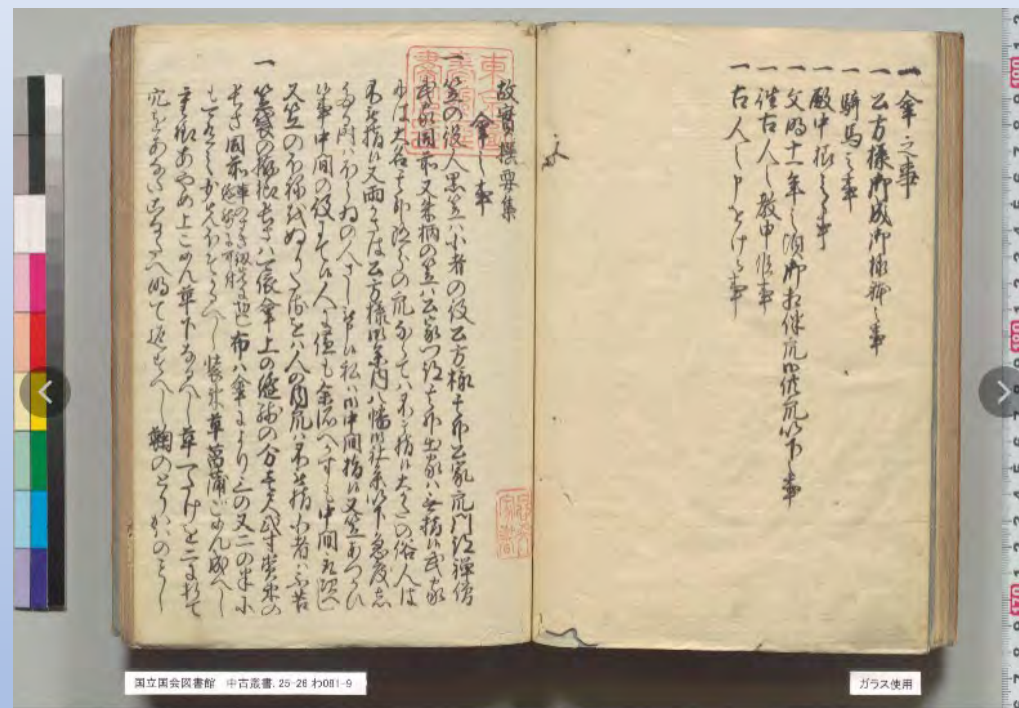
間紙の例

<https://dl.ndl.go.jp/pid/11891802/1/74>

間紙なし



間紙あり



折込図の例

<https://dl.ndl.go.jp/pid/11891876/1/46>



4-3. 業者決定（入札）

実施決裁→入札公告→入札説明会

→提案書確認→一般競争入札

- ・ 入札説明会で、仕様書の説明
- ・ 提案書を精査し、入札参加業者の審査
- ・ 落札業者の一覧はＨＰで公開

<https://www.ndl.go.jp/jp/supply/keiyaku/rakusatsu/index.html>

4-4. データ作成（受託者作業）

【館内作業】（毎日の作業）

- (1) 撮影資料引き渡し
 - ・ 当館と受託者で資料名・数量リストを照合
 - ・ 疑義照会
- (2) 資料に付着した埃の掃除・資料確認
- (3) 撮影
- (4) 撮影資料返却（リスト照合・疑義照会）

【館外作業】

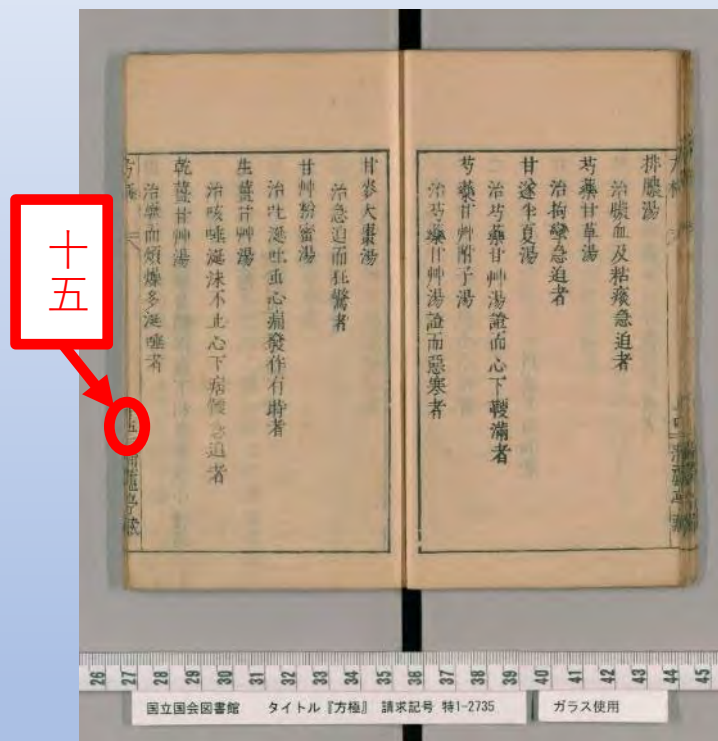
- (5) 画像データ作成（トリミング等）
- (6) メタデータ・管理データ等作成

4-5. 納品・検収

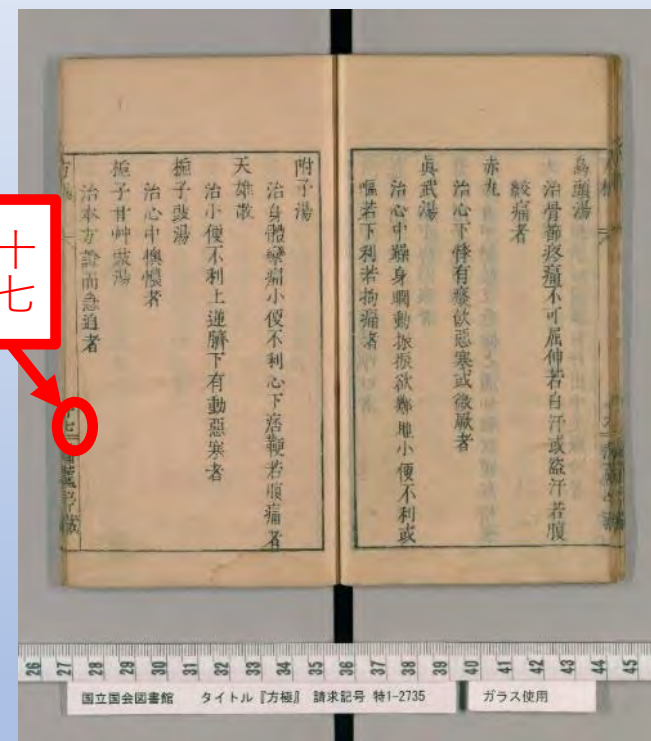
数回に分けて納品 → 職員による検収



画像データ不備の例：撮影漏れ（丁付けから発見可能）



<1コマ目>



<2コマ目>

この間に
撮影漏れあり

画像データ不備の例：撮影漏れ（画像だけでは発見不能）



<1コマ目>

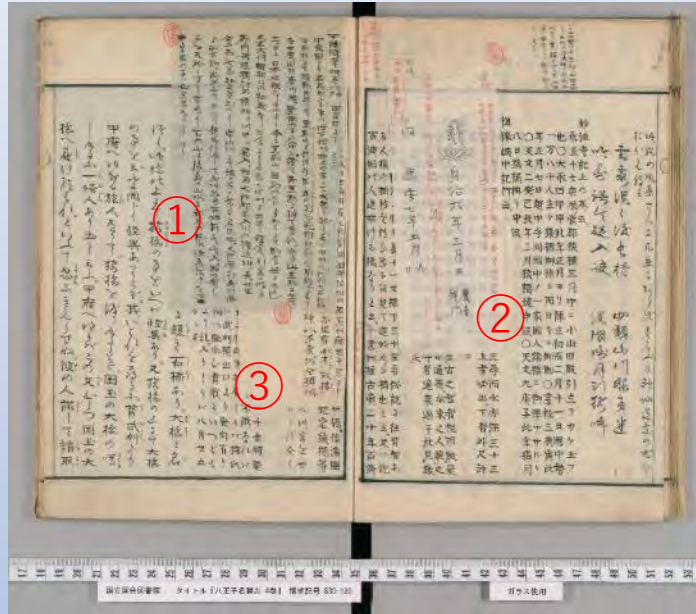
この間に
撮影漏れあり



<2コマ目>

画像データ不備の例：付箋の上げ漏れ

この間に、③の付箋を上げた画像が必要



<1コマ目>
付箋が3枚下りている



<2コマ目>
①②の付箋が上がる



<3コマ目> 次頁

画像データ不備の例：名票の誤り（タイトル）



〈参考〉 綴じ糸を切った資料の仮止め

タイトル『牽牛花集』 請求記号 特7-614

正しくは『牽牛花目録』

4-6. 画像とデータの登録・公開（1）

- ・画像データ、メタデータ、管理データを「国立国会図書館デジタルコレクション」に登録
- ・デジタルコレクションとHPにお知らせ掲載

2024年3月28日 図書、雑誌、古典籍等約19.9万点を「国立国会図書館デジタルコレクション」に追加しました

国立国会図書館は、次の約19.9万点を「[国立国会図書館デジタルコレクション](#)」に追加しました。

コレクション	点数	公開範囲
図書	約192,300点	館内限定 (一部インターネット公開)
雑誌	約500点	館内限定 (一部図書館向け・個人向け送信)
古典籍	約3,200点	館内限定 (一部インターネット公開)
日本占領関係資料	約800点	インターネット公開 (一部館内限定)

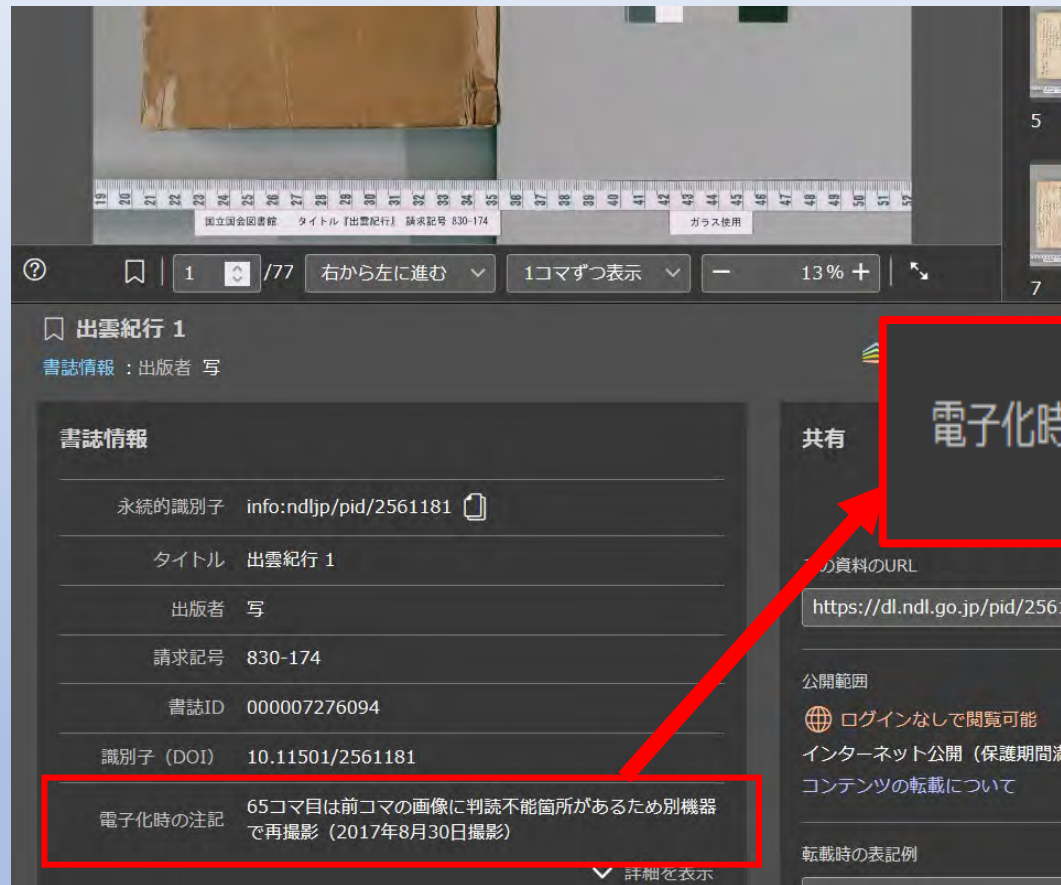
4-6. 画像とデータの登録・公開 (2)

公開後のメンテナンス

- ・ 画像データに不備がある場合の訂正
(受託者による対応)
- ・ 書誌事項等に不備がある場合の訂正
(職員による対応)
- ・ 解題の登録 (一部の資料)
 <解題一覧> <https://dl.ndl.go.jp/ja/kaidai.html>
 <解題のある資料の例>
 <https://dl.ndl.go.jp/pid/8929276>

公開後に画像データ不備が発見された例（再撮影）

<https://dl.ndl.go.jp/pid/2561181/1/65>

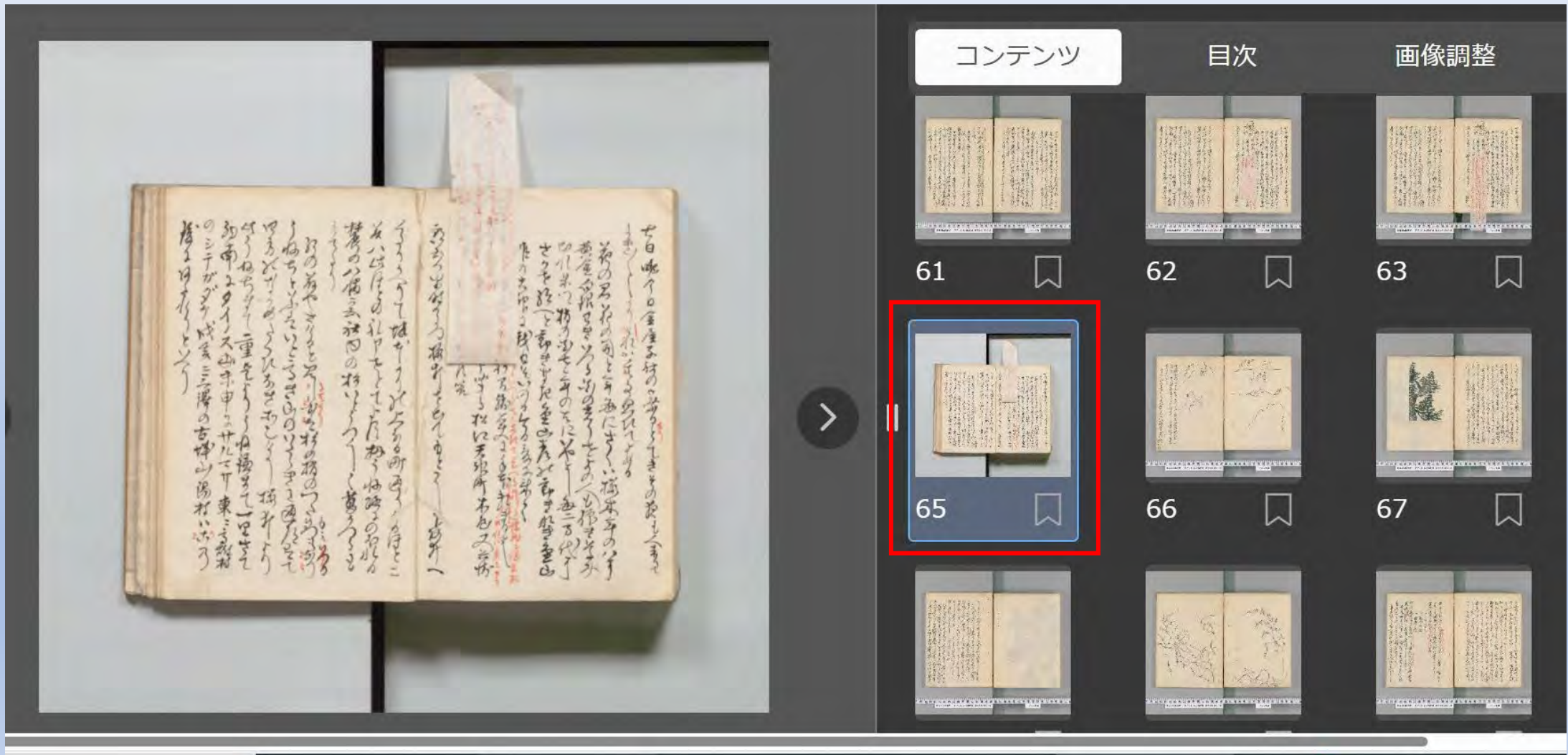


電子化時の注記

65コマ目は前コマの画像に判読不能箇所があるため別機器で再撮影（2017年8月30日撮影）

公開後に画像データ不備が発見された例（再撮影）

<https://dl.ndl.go.jp/pid/2561181/1/65>



公開後に画像データ不備が発見された例（暫定措置）



電子化時の注記 31コマ目と32コマ目の間に1コマ分未撮影箇所あり

書誌情報

永続的識別子 [info:ndljp/pid/2566706](https://nii.jp/pid/2566706)

請求記号 854-21

書誌ID 000007726280

識別子 (DOI) 10.11501/2566706

電子化時の注記 31コマ目と32コマ目の間に1コマ分未撮影箇所あり

▼ 詳細を表示

5. 内製デジタル化の工程

1. 対象資料選定

2. 作業マニュアルの作成

~~業者決定(入札)~~

3. データ作成

メタデータ⇒撮影⇒画像データ作成

4. データ検収

5. 画像とデータの登録・公開

5-1. 対象資料選定

①対象資料の選定

貴重書・準貴重書以外の和古書・漢籍

⇒近代和装本・写本から選定

2021年度は1,900コマ（2021年11月から開始）

2022年度は45,000コマ

2023年度は29,000コマ



②現物の状態確認

※外注で必要だったコマ数の確認は不要



③対象資料決定・リスト作成

5-2. 作業マニュアルの作成

撮影場所…国立国会図書館内

撮影作業員…職員

撮影方法…館内共有のスキャナにより平置き資料を垂直上方から撮影

資料の取り扱い…原則ガラス使用

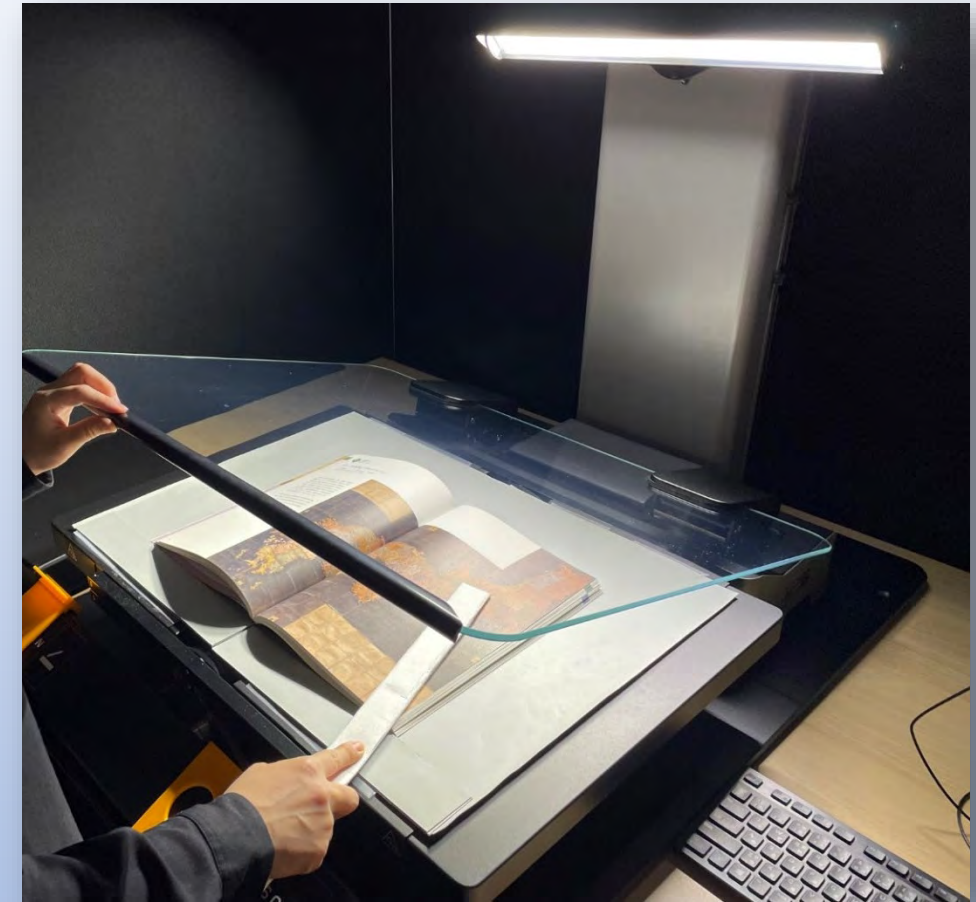
写込み物…最初の一コマ目のみカラーチャート・メジャー、名票類は入れない

撮影順…分割・付箋・折込・挟み込み物等

画像のフォーマット…撮影画像をTIFFで保存し、その後、夜間の自動処理により、JPEG2000形式に変換して保存する。

5-3.データ作成

- ① 管理担当職員がメタデータ投入
- ② 撮影資料受け渡し
- ③ 撮影担当職員が資料に付着した
埃の掃除・資料確認
- ④ 撮影担当職員が機器を準備
- ⑤ 撮影担当職員が撮影
- ⑥ 資料返却（リスト照合・疑問点の確認）
- ⑦ 画像データ作成



検収でエラーが確認された資料について、同様に再撮影

5-4.データ検収

- ① 撮影担当職員とは別の検収担当職員により検収。必要に応じて、原本と比較しながら全コマを確認
- ② 管理担当職員は再撮影が必要なら差し戻す
- ③ 撮影担当職員が、エラー内容に即して該当のコマを再撮影
- ④ 検収担当職員が再撮影したコマを再度検収
- ⑤ ①～④を繰り返し、エラーがなくなったら、管理担当職員が最終確認。問題なければ、ステータスを完了に変更
- ⑥ 資料を書庫に納架し、進捗を記録

5-5.画像とデータの登録・公開

データ登録

- ・提供用のシステムにデータを登録し、公開

公開後のメンテナンス

- ・画像データに不備がある場合の修正
(職員による再撮影及び画像差し替え)
- ・書誌事項等に不備がある場合の訂正
(職員による修正)

6. マイクロ資料からのデジタル化（1）

- ・ 外注によるデジタル化（2022年度）

近代の新聞・雑誌のうち、準貴重書を撮影したマイクロフィッシュ
630枚（約27,000コマ）

『共存雑誌』『新思潮』『中外新聞』『有喜世新聞』など116タイトル

< メリット >

原資料に負荷がかからない

コストが原資料からのデジタル化に比べて低い

準備作業が原資料からのデジタル化に比べて少ない

資料を館外持ち出し可とし、館外での撮影作業を可能としたため、館内に
撮影場所の確保が不要

< デメリット >

画質が原資料からのデジタル化に比べて悪い

6. マイクロ資料のデジタル化（2）

- ・ 内製によるデジタル化

当館作製マイクロフィルムから

2023年度：400リール≒20万コマ

< メリット >

短時間で多くの資料をデジタル化できる

準備作業が原資料からのデジタル化に比べて少ない

< デメリット >

画質が原資料からのデジタル化に比べて悪い

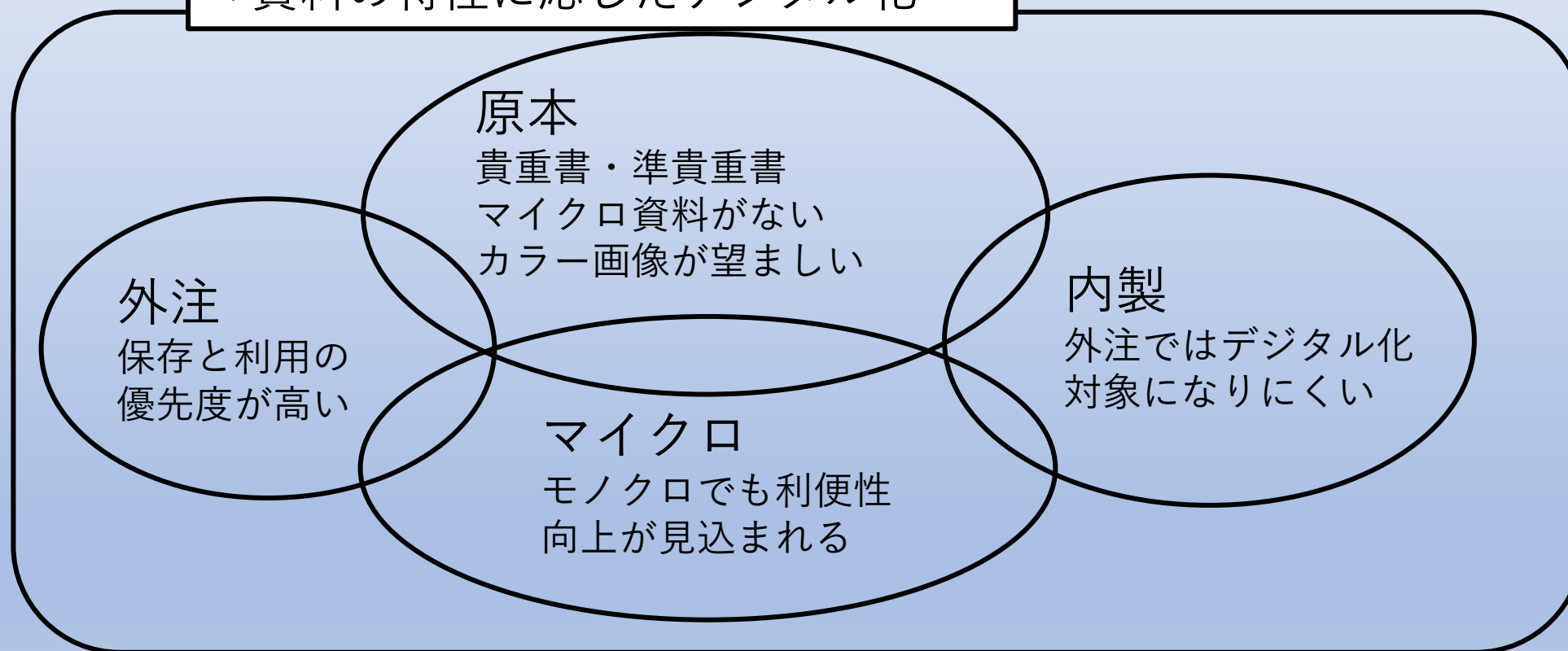
→ マイクロからのデジタル化に適さないもの（彩色資料など）は、
原本からのデジタル化に切り替え



7.まとめ


できるだけ早く、多く、資料を傷めずに

⇒資料の特性に応じたデジタル化



参考：国立国会図書館デジタルコレクションの機能（1）

<https://dl.ndl.go.jp/>
<https://dl.ndl.go.jp/help#idx3-2>



The screenshot shows the National Diet Library Digital Collection website. The main search area is titled '古典籍資料（貴重書等）' (Classical Literature (Valuable Books)). It displays a search result for '大津（月岡）' (Otsu (Gyokko)) with 82108 items, 17234 items, and 7 items. The search filters are set to 'Title', 'Author', and 'Publisher'. The search results are displayed in a grid format, showing various items with their titles and authors.

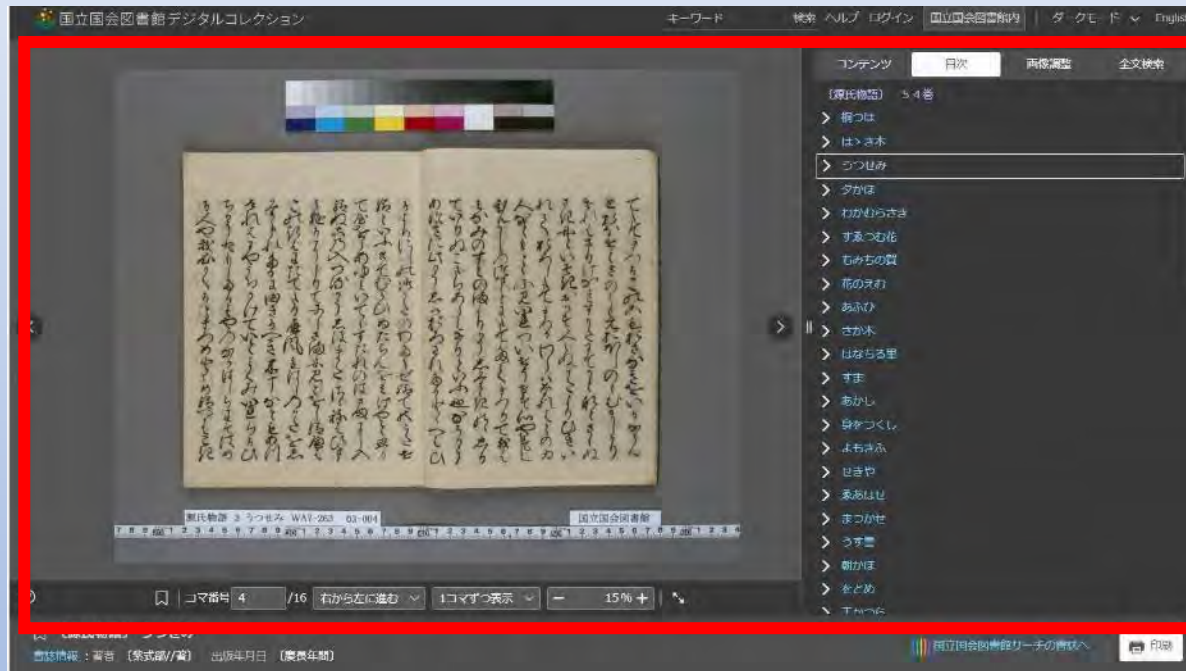
3-2 ファセットナビゲーション
 以下の項目を組み合わせて検索結果を絞り込み、検索結果の表示を変更することができます。
 ※現在、国立国会図書館内の利用者端末でMozilla Firefoxを利用して閲覧した場合、ファセットナビゲーションが使用できないことが確認されています。今後対応予定です。もう一度ファセットの項目を選択すると絞り込みを行うことができますので、お試しください。

ファセット項目	説明
検索する項目	<ul style="list-style-type: none"> 「書誌情報」：タイトル等のメタデータを対象に検索します。目次・本文は検索対象に含まれません。 「資料の本文」：資料の本文のみを対象に検索します。タイトルやメタデータは検索対象に含まれません。 「資料の目次」：目次を対象に検索します。
閲覧方法	資料の公開範囲によって絞り込みができます。各公開範囲の詳細については「よくあるご質問」「デジタル化資料の印刷や保存はできないのですか。」をご覧ください。
デジタル化資料	「デジタル化資料」を頂点としたコレクション体系の中から特定の集合を絞り込むことができます。
電子書籍・電子雑誌	「電子書籍・電子雑誌」を頂点としたコレクション体系の中から特定の集合を絞り込むことができます。
NDC分類	NDC（日本十進分類法）によって絞り込みができます。
出版年	出版年で検索結果を絞り込むことができます。
歴史的音源ジャンル	コレクション「歴史的音源」専用の項目です。
収集根拠	コンテンツの収集についての情報を記載しています。
検索結果をまとめる	「同一タイトルでまとめる」「巻号単位でまとめる」「記事単位のみ表示」から選んで、表示数を絞ることができます。 <ul style="list-style-type: none"> タイトルでまとめる → 雑誌のタイトルに検索語の文字列がある資料のみ（全号まとめのみ）表示されます。雑誌の記事タイトルにのみ検索語が含まれる資料は、表示されません。 巻号でまとめる → 検索語の文字列がある資料の巻号のリストを表示します。 記事単位のみ表示 → 電子書籍・電子雑誌が対象です。記事単位で表示します。
官報種別	コレクション「官報」専用の項目です。
記事種別	コレクション「官報」専用の項目です。
機関	コレクション「官報」専用の項目です。
件名	コレクション「電子書籍・電子雑誌」専用の項目です。件名によって絞り込みができます。
古典籍分類	コレクション「古典籍」専用の項目です。国書院目録の分類、四部分類、絵図／地域名（旧国名・都市名）によって絞り込みができます。

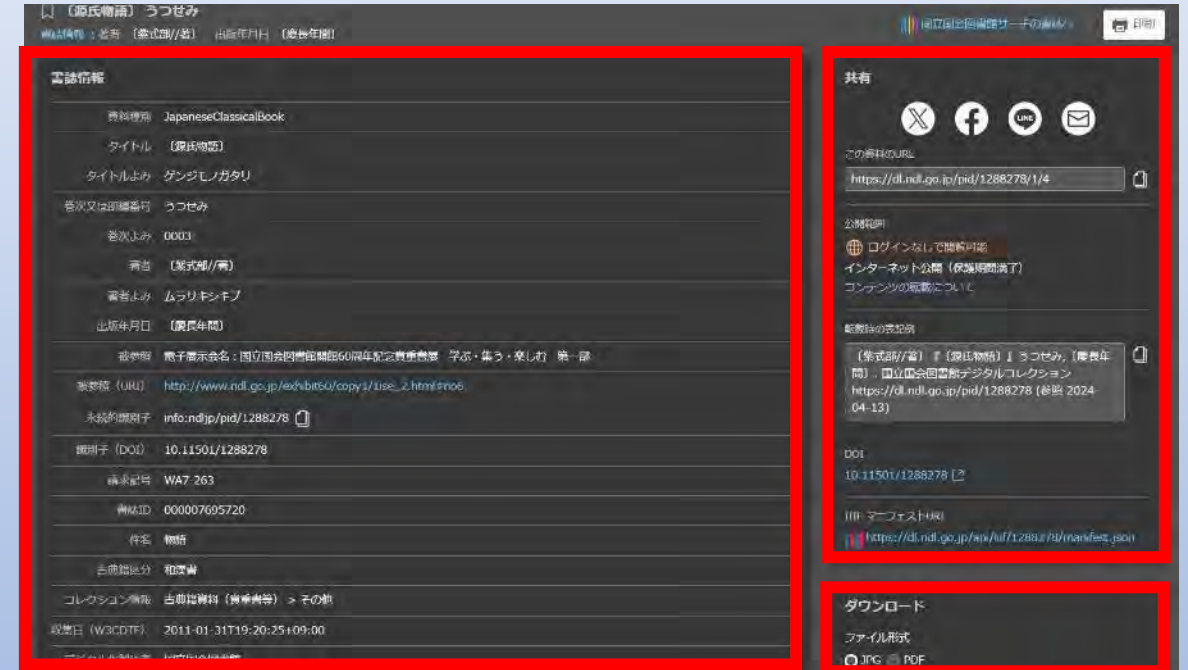
「コレクション古典籍資料（貴重書等）」⇒「すべての条件」⇒古典籍資料専用の検索画面
 錦絵や絵図、漢籍、印刷資料や書写資料など、資料の種類で絞り込んだ検索も可能

参考：国立国会図書館デジタルコレクションの機能（2-1）

<https://dl.ndl.go.jp/pid/1288278>



コンテンツ表示エリア

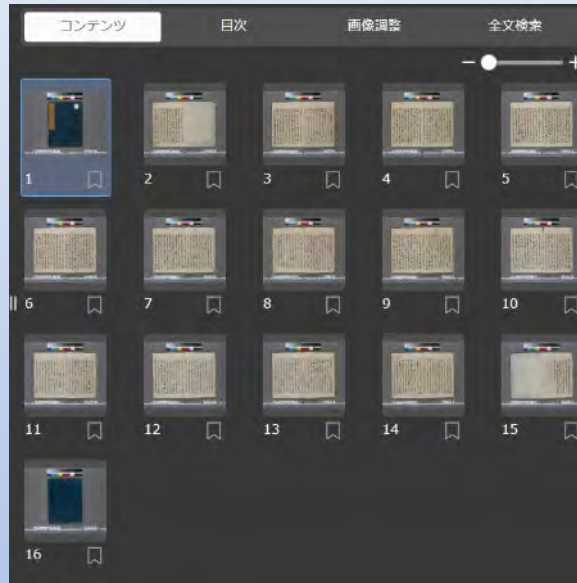


書誌情報エリア

共有エリア

ダウンロードエリア

参考：国立国会図書館デジタルコレクションの機能（2-2）



<コンテンツ>タブ
収録されている画像の
サムネイルを一覧する
ことができる

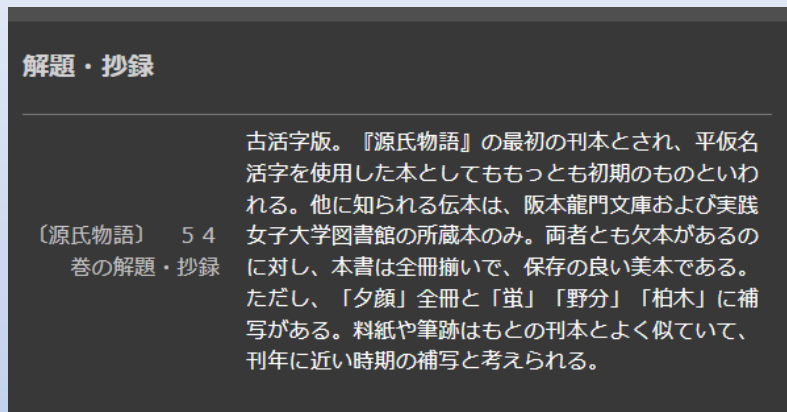


<目次>タブ
階層構造で目次・巻号を
見ることができる



<画質調整>
デジタル化画像の表示を
調整できる

参考：国立国会図書館デジタルコレクションの機能（2-3）

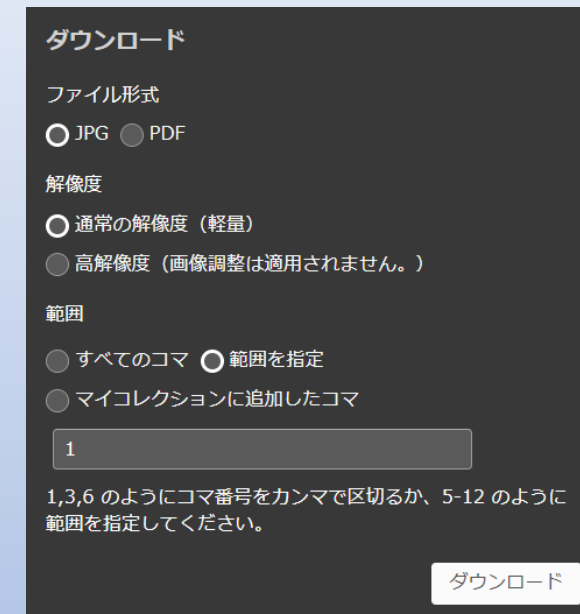


解説・抄録

一部資料の書誌情報エリアにある

共有エリア

SNSで情報を共有するためのボタンやIIIFマニフェストURIなどがある



ダウンロードエリア

ファイル形式、サイズ、コマ範囲を指定する

参考：次世代デジタルライブラリー

<https://lab.ndl.go.jp/service/tsugidigi/>



NDL Lab
National Diet Library, Japan

[本文へ](#) [サイトマップ](#) [About us \(English\)](#) [国立国会図書館ホームへ](#)

ホーム

サービス (体験する)

データ (活用する)

イベント (参加する)

NDLラボについて

ホーム > サービス (体験する) > 次世代デジタルライブラリー

サービス (体験する)

次世代デジタルライブラリー

Next Digital Library (English)

NDL Ngram Viewer

NDC Predictor

ジャパンサーチ

書誌情報検索・可視化システム

翻デジ

国デコ Image Wall

電子読書支援システム

次世代デジタルライブラリー

サービスURL

<https://lab.ndl.go.jp/dl/>

ソースコード

<https://github.com/ndl-lab/tsugidigi-web/>

概要

次世代デジタルライブラリーは、国立国会図書館次世代システム開発研究室での研究を基に開発した機能を実装した実験的な検索サービスです。全文テキスト検索機能や機械学習を用いた自動処理、International Image Interoperability Framework (IIIF) API等の技術的有効性を検証することを目的としています。

検索対象は「[国立国会図書館デジタルコレクション](#)」でインターネット公開をしている著作権保護期間満了となった図書及び古典籍のデジタル化資料（約35万点）です。

「全文から検索する」では全文テキストからの検索ができます。令和5年1月現在の対象資料は、次のとおりです。

- 令和3年度デジタル化資料のOCRテキスト化事業の成果物を利用した図書資料約28万点
- 令和4年度古典籍資料のOCRテキスト化実験の成果物である古典籍資料約8万点

「画像から検索する」では図書及び古典籍資料から自動的に抽出された画像・図版の中から、似たものを探すことができます。

https://lab.ndl.go.jp/data_set/r4ocr/r4_koten/

古典籍資料のOCRテキスト化実験

令和4年度、[令和3年度OCR処理プログラム研究開発](#)等を得た知見を生かし、古典籍資料を対象としたOCR処理プログラム（以下、「NDL古典籍OCR」といいます。）の内製開発を実施し、古典籍資料のテキスト化実験を行いました。

1. 古典籍資料のOCRテキスト化実験の目的

昨今のAI（機械学習）を用いたOCR（光学的文字認識）処理技術の進展により、画像データからテキストデータを作成し、本文検索サービスとして提供することが可能となりました。当館においても、[令和3年度OCR関連事業](#)において、当館が提供するほぼ全てのデジタル化資料をOCR処理によってテキスト化し、デジタル化資料の本文検索サービスの実現に向けて取り組んでいるところです。

[国立国会図書館デジタルコレクション](#)から提供している古典籍資料は、[令和3年度OCRテキスト化作業](#)の対象外でした。

こうした資料の多くには、くずし字や異体字、変体仮名等が使われており、専門的な知識がないと判読自体が難しい資料が多く存在します。

利用者が必要とする古典籍資料をより簡単に発見できるようになることは、単に利便性の向上のみならず、パブリックドメインとなった資料の利活用促進の観点からも大きな意義があります。

そのため、まずは本文を文字列として検索できるようにし、求める資料にたどり着きやすくすることが重要です。

古典籍資料を対象としたOCRテキスト化の先行事例としては、人文学オープンデータ共同利用センターが開発している『[みを](#)』や、凸版印刷株式会社が開発している『[ふみのほ](#)』といったくずし字OCRアプリケーションが存在しますが、これらはスマートフォンやタブレット端末の専用アプリケーションとして、利用者が撮影した個々の古典籍資料の判読のサポートやテキスト化を行える機能を備えたものです。

一方で、当館のような数百万画像規模の古典籍資料を所有する機関がOCRによるテキスト化を実施するためには、自らの計算機サーバ上で大量の画像に対して高速なテキスト化処理を行えるOCR処理プログラムが必要となります。