令和7年8月7日(木) 令和7年度東北地区大学図書館協議会合同研修会 事例発表2

日本古典籍のデジタルアーカイブ構築と国内外の所蔵機関との連携

― 国文学研究資料館の取り組み

国文学研究資料館 学術情報課 基盤データ企画係 山下 日向子

国文学研究資料館ってどんなところ?





国文学研究資料館とは

日本文学と関連資料を集積・提供する総合研究機関として、国内外の研究者と連携しながら、古典籍*を活用した分野横断的な共同研究を推進しています。50年以上の資料研究の蓄積を基に、日本文学研究の基盤強化と知的資源の活用を図っています。

■ 基本情報

大学共同利用機関法人 人間文化研究機構 国文学研究資料館

・教員:27名、事務系職員:44名

・所在:東京都立川市

個別の大学では維持が困難な大規模設備や膨大な資料・情報等を提供し、国内外の研究者による効果的な共同研究を支援する研究機関

■ 主な沿革

*主に江戸時代以前の書籍

1972年5月 国文学研究資料館 創設(品川区) ⇒文部省史料館(1951年設置)を組み入れ

2003年4月 総研大日本文学研究専攻 設置

2008年3月 立川市へ移転

2014年4月 フロンティア事業 開始

(歴史的典籍NW事業)

2024年4月 フロンティア事業 継続開始 (DDHプロジェクト)

国文研DDHプロジェクトについて



データ駆動による課題解決型人文学の創成 〜データ基盤の構築・活用による次世代型人文学研究の開拓〜 (略称・国文研DDHプロジェクト)

- ■文系唯一の大規模学術フロンティア促進事業 (文部科学省)
- ■2024年度から2033年度までの10年計画で実施
- ■目指すもの

日本文学を中心とするデータインフラを構築し、さまざまな課題意識に基づく国内外・異分野の研究者との協働による大規模データを活用した次世代型人文学研究を開拓すること

国内外機関等との 連携による 更なる画像データの拡充

画像データのAI利活用等 によるテキストデータ化

データ分析技術開発の推進

国文研DDHプロジェクトの概要(2つの柱)

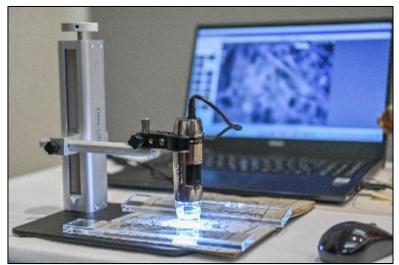


- 研究基盤データベースのデータ拡充・高度化
- ・画像データの集積の範囲を 近代初頭(明治時代初頭)まで拡張
- ・機械可読の大規模テキストデータを形成
- 異分野融合によるデータ駆動型研究の推進

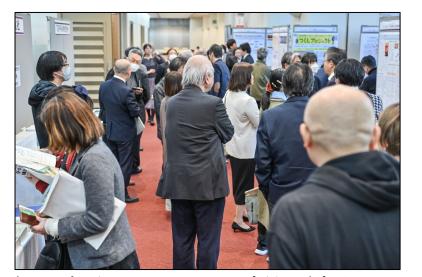
情報学、自然科学分野などとの協働による データ駆動型研究や基盤技術の高度化を推進







基盤技術の高度化の一例(書物のマテリアル分析・解析)



国文研DDHプロジェクトキックオフシンポジウム(ポスターセッション)

国書データベースについて



国書データベース



https://kokusho.nijl.ac.jp/
★ぜひアクセスしてみてください!

国内外の機関や個人が所蔵する書誌約95万点、デジタル画像約30万点を無料で公開する世界最大級の日本古典籍データベースです。





『浦しま』(国文学研究資料館所蔵)出典: 国書データベース, https://doi.org/10.20730/200017771

特徴その1

探しやすい

タイトル キーワード 画像に付けられたタグ 本文テキスト ピックアップコンテンツなど…様々な条件で検索できます!

特徴その2

引用しやすい



電子データに付与される国際的識別子、DOI(デジタルオブジェクト識別子)を採用!リンク切れを防ぎます。

特徴その3

活用しやすい



IIIF(International Image Interoperability Framework)に対応したビューアを使用。異なるデジタルアーカイブ間での画像共有や再利用が容易です。



国書データベース

https://kokusho.nijl.ac.jp/
★ぜひアクセスしてみてください!



1972年 国文学研究資料館創設

2006年 「日本古典籍総合目録データベース」を公開

2014年 大規模学術フロンティア促進事業

「日本語の歴史的典籍の国際共同研究ネットワーク構築計画」開始

2017年 「新日本古典籍総合データベース」を公開

2023年 「日本古典籍総合目録データベース」と

「新日本古典籍総合データベース」を統合し、

「国書データベース」を公開

「コーニツキー版欧州所在日本古典籍総合目録」を統合

2024年 大規模学術フロンティア促進事業

「データ駆動による課題解決型人文学の創成」開始

「近代書誌・近代画像データベース」統合

国書データベースの使い方

こちらを ご参照ください!

- ① 国書データベースの使い方 https://kokusho.nijl.ac.jp/page/guide.html
- ② 国書データベース パンフレット https://kokusho.nijl.ac.jp/page/kokushopamphlet 202309 web JP.pdf
- ③ 国書データベースの検索方法 (日本古典籍講習会資料) https://kokubunken.repo.nii.ac.jp/records/2000398







「使いやすい!」を目指して①

利用手続きのロゴ表示や各機関の利用条件一覧の作成を行い、利用条件がすぐに分かるように工夫。 利用条件のパブリックドメイン化を推進するなど、よりオープンな利用を促進しています。

★利用手続きのロゴ表示

- ・画像ビューアの下に利用条件を表す**ロゴを表示、利用方法詳細へもアクセス**可能
- ・国際的枠組みに対応した条件を使用
- →どのような手続きが必要か一目で分かるように!

○近年、著作権保護期間"満了"の資料に推奨

- パブリックドメイン(制限なし)
- ・Rights Statements(所蔵機関が条件を明示)

国文学研究資料館は2024年6月に 所蔵資料26,000点を パブリックドメインに変更!







「使いやすい!」を目指して②



★利用条件一覧の作成



画像利用条件(他機関等所蔵)

1. 他機関等所蔵資料の利用 (国文研所蔵資料の利用条件については<u>こちら)</u>

国書データベースで公開されている画像の利用条件は原資料の所蔵者ごとに異なります。

利用条件が示す内容について 確認できるサイトヘリンク

※パブリックドメインの公式説明ページ等

到 0015 九州大学附属図書館 細川文庫 206 ム <u>一覧へ</u> <u>プミ内へ</u> <u>パブリックドメイン</u> にて	て公開
団 0250 九州大学附属図書館 支子文庫 等 233 <u>マイクロフィル ● 画像 ● 案内へ パブリックドメイン</u> にて <u>一覧へ</u>	て公開
団 0047 九州大学文学部 598 ス <u>マイクロフィル ● 画像 シ素内へ デジへ プリックドメイン</u> にて <u>一覧へ</u> フジへ	て公開
団 KYUS 九州大学附属図書館(医学図書館) 946 デジタ <u>● 画像 デジタ アンドメイン</u> にて	て公開
B KYUS 九州大学附属図書館(中央図書館) タル ● 画像 <u>一覧へ</u> パブリックドメインにて	て公開
崎 0395 諏訪神社 201 マイクロフィル <mark>→ 画像 → 案内へ </mark>	o Copyright - Contractual
場 SUWJ 諏訪神社 50 デジタル ② <u>画像</u> ② <u>案内へ</u> Rights Statements No ® <u>Pestrictions</u> にて公開	o Copyright - Contractual

各機関の公開画像一覧 ヘリンク

各機関の利用案内ページ 等ヘリンク



※利用条件や手続きに関するウェブページ

「使いやすい!」を目指して③



画像利活用の促進 画像からテキストへの変換

国立国会図書館等との連携による、AIを活用した、OCR/機械処理等による全文テキスト化を実施(目標点数27万点)

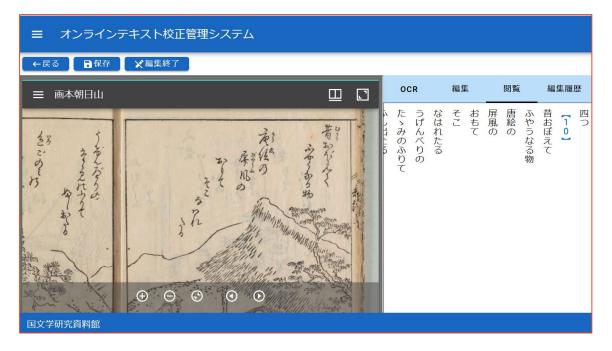
⇒認識率の向上/利活用(検索)へ

チェックを経たテキストの公開 (現代語訳なども含む)等

⇒人力依存からの省力化を目指す研究 と実装の連動

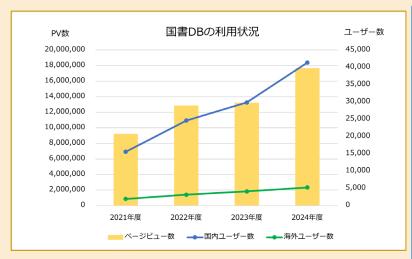
機械処理データからの校正管理システム (研究者による開発)

- ⇒進捗管理上のデータ収集
- ⇒改良および実装に向けた取組へ



- ・ツール改良により大学等での演習活用を可能に
- ・利用の幅を拡げ、国際的利用も視野に

国書データベースの利用状況



	ナる利用状況 ヶ月〉の状況
【主な	以内訳】
中国	1,375人
米国	877人
台湾地域	265人
韓国	249人
香港	181人
仏国	145人
英国	137人
独国	95人
ロシア	57人
シンガポー	-ル 56人

年度	ページビュー数	国内ユーザー数 月平均	海外ユーザー数 月平均
2020年度	7, 104, 885	12,809	1, 221
2021年度	9, 225, 119	15,560	1, 940
2022年度	12, 869, 292	24,608	3, 162
2023年度	13, 241, 736	29,829	4, 096
2024年度	17, 676, 757	41, 376	5, 240

2024年度にページビュー数 1700万件突破 さらに拡大中

対象機関の拡大



博物館・研究所など

新たな資料所蔵機関と交渉を重ね、対象機関を拡大

所蔵機関数(個人含) 国内外 約700機関

大英図書館 韓国国立中央図書館 フリーア美術館 など アジア・欧米の海外機関も公開





広く認知された結果 「静嘉堂文庫」をはじめ、所蔵機関から 本プロジェクトへの参画を要望する申し出が 多数寄せられた



国内外に残る古典籍の公開に取組む世界に開かれたポータルサイトへ

つながる 他機関データベースとの連携等





海外ポータルサイト等との連携

Web NDL Authorities

国立国会図書館典拠データ検索・提供サービス

NDL Authorities (国立国会図書館)



国立情報学研究所(NII)

Japan Knowledge

国際的に利用されている 商用データベースと連携

人文学オープンデータ共同利用センター Center for Open Data in the Humanities

№ / 日本古典籍データセット

ERDB-JP Electronic Resources Database-JAPAN

国書データベースのデータ拡充

Peter Kornicki先生作成 「欧州 (Europe) 所在日本古書総合目録」 書誌データを統合

画像の作成から公開まで

画像作成から国書データベース公開まで



業者撮影 撮影専門業者

メリット

- ・さまざまな資料に対応可能
- ・状態が悪い資料も撮影可能

デジタルカメラを用いて撮影

デメリット

- ・契約(撮影枚数の確定)が手間
- ・所蔵者と業者との日程調整が必要
- ・経費(撮影単価)が高い



画像解像度:400dpi 1日6時間500~600枚 撮影可

内 製 スキャナーを設置し、機関職員等が撮影

メリット

- ・資料の持ち出しをせずに機関内で撮影可能
- ・事前の撮影枚数の確定が不要
- ・時期を問わず(通年)撮影が可能

①国文研内製:所蔵資料を借用し、当館

スタッフが設置スキャナにて撮影

②機関内製:所蔵者にスキャナを

貸し出して現地スタッフで撮影





マイクロコンバート



画像提供

画像形式変換等





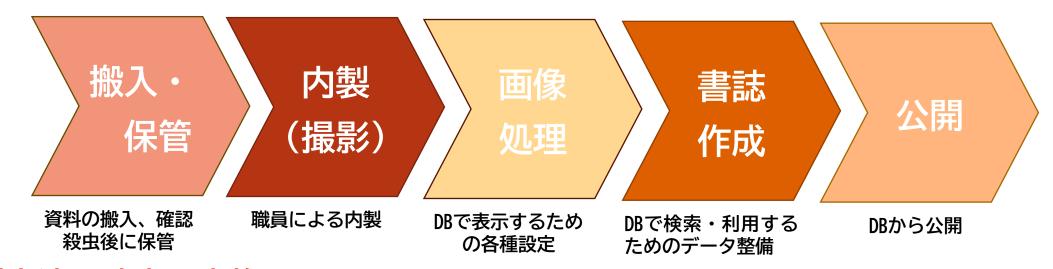


書誌データ付与



画像作成から国書データベース公開まで





>内製方法の確立と実施

2014(平成26)年度から試行を重ね、その方法を確立し、2018(平成30)年度から本格的に実施

> 内製実施機関の例

連携機関との協力体制の構築

◆福島大学

◆九州大学

- ◆江戸東京博物館 ◆愛媛大学
- ◆加賀市立図書館
- ◆慶應義塾大学
- ◆研医会
- ◆千葉県立中央博物館◆鬪雞神社
- ▼十朵宗立中央傳物路▼飼雞仲仏
- ◆東京書籍附設教科書図書館東書文庫

- ◆宮城教育大学・専修大学
- ◆大谷大学 ◆帯広畜産大学
- ◆京都大学 ◆共立女子大学
- ◆清光山西嚴寺◆東京藝術大学
- ◆日本体育大学◆野中烏犀園
 - ◆嵐牛俳諧資料館

			内製点数			
~2018 (H30) 【試行期間】	2019 (R1)	2020 (R2)	2021 (R3)	2022 (R4)	2023 (R5)	2024 (R6)
1,968	4,454	5,268	3,928	6,985	8,530	8,315

内製機関数							
	2018 (H30)						
2	9	9	14	17	22	24	22

書誌作成

五文学研究資料館 国文学研究資料館

画像をデータベースで検索・利用する上で必要な「書名/著者名/所蔵者」などの様々な書誌データ(メタデータ)を整備する



くずし字の読解可能な職員が、資料画像から書誌データを作成し、入力していく。

*6時間で15点程度の書誌情報作成可能

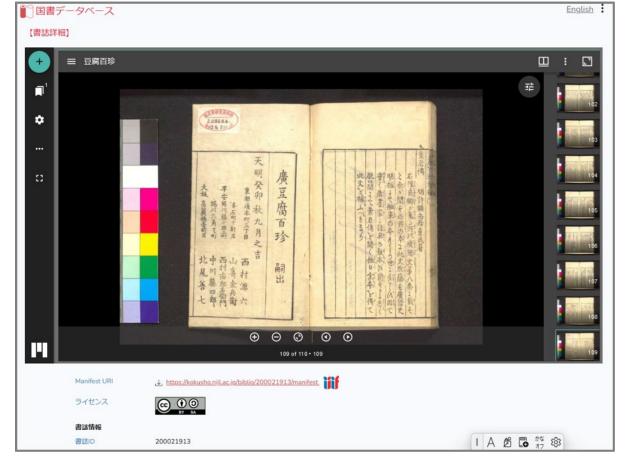


公開・保存

国文学研究資料館

画像と書誌とをリンクさせ公開

撮影画像を変換し公開用画像を作成。基準を満たさない画像がないか機械的に1枚ずつ確認し、書誌データとともにデータベースに取り込んでいく



データ保管とバックアップ



- ・公開画像は複数のバックアップ体制⇒国書データベースのストレージやクラウド
- ・オリジナル画像は光ディスク等複数媒体で保管 ⇒永年保存には予算・人員が課題(NDL等も同様) ⇒所蔵機関にはコピーを送付
- ・ 画像データの保管と運用を定めた 『デジタル保存ポリシー』の策定を検討中

国内外の所蔵機関との連携強化に向けた今後の取組



メタデータの標準化

「メタデータ流通ガイドライン」

国立国会図書館(NDL)、オープンアクセスリポジトリ推進協会(JPCOAR)、 これからの学術情報システム構築検討委員会、NIIが共同で作成。 コンテンツの発見可能性を高めるための持続可能性のあるメタデータ流通の実現を目指す。

2023年6月にメタデータ流通ガイドライン「古典籍編」を公開

- →国文学研究資料館も作成に協力!
- ・古典籍(貴重書・準貴重書、江戸期以前の和漢書など)を対象としたメタデータ設計および流通に関するガイドライン

今後の検討課題

- ・「漢籍」「洋古書」等への範囲拡大
- ・古典籍特有の事項をどこまで項目として切り出すか
- ・メタデータ階層は著作単位か個別資料単位か



メタデータ流通ガイドライン(古典籍編)

https://ndlsearch.ndl.go.jp/guideline/historical



外部サービスとの互換性確保

対応項目	概要・対応内容	備考・関連リンク
JPCOARスキーマや DC-NDL	・国内連携機関での機関リポジトリからの公開支援 ・Japan Search との連携	https://jpsearch.go.jp/cooperation on →Japan Search 「連携をご 希望の機関の方へ」
NACSIS-CAT 書誌登録のサポート	・CATPマッピング(登録機関の目録公開サポート) ・著作単位の記述等、NCR2018適用細則に対応	https://contents.nii.ac.jp/korek ara/about/sw_wg/pc202202 →NCR2018適用細則
TEI(Text Encoding Initiative) ガイドライン	・翻刻・解題など様々な付加情報の記述に対応可能 → デジタル人文学からの要請	https://github.com/TEI-EAJ/jp_guidelines/wiki→日本語向けTEIガイドライン
国際的枠組みとの協調	・国際的に使用されるメタデータスキーマ・枠組み との協調 MARC21、MODS、DARIAH-EUなど	「これからの学術情報システ ム構築検討委員会」での検討

その他

①国書データベースと所蔵機関双方から画像公開する場合のガイドライン 作成

所蔵機関による自機関のデジタル公開サイトが構築され、国書データベースと同一資料の画像が 別途公開される事例が増加。

⇒ジャパンサーチ等のポータルサイトとの連携を進める上で、利用者がポータルサイト上で 混乱なく利活用できるよう、識別や表示の仕組みの整備を検討中。

②デジタルアーカイブのメタデータ流通に関係する委員会への参画

学術情報流通に関係する団体・委員会へ職員が参加。

当館データベースの**メタデータ標準化・高度化とともに、国内の古典籍所蔵機関とのメタデータに** 関する連携強化を図る。

- 1) オープンアクセスリポジトリ推進協会(JPCOAR)
- 2) これからの学術情報システム構築検討委員会

③コミュニティサービスの構築

連携機関等と情報共有やデータの利用状況等を共有するための仕組みを検討し、情報提供プラットフォームを構築することを目指す。

国書データベースと 連携してみません か?

古典籍目録の掲載だけでも大歓迎! (紙・電子いずれもOK)

古典籍画像と書誌情報があれば公開可能!

自機関ウェブサイトでの公開も もちろん継続可能

画像公開を計画している段階でも ぜひご相談ください!

連絡先:国文学研究資料館

基盤データ企画係

Mail: center_db@nijl.ac.jp

まずはお気軽に ご相談ください! ご連絡をお待ちし ております。

ご清聴ありがとうございました!

1 研究基盤データベースのデータ拡充・高度化

(参考資料) 国文学研究資料館

画像データの集積の範囲を近代初頭(明治時代初頭)にまで拡張するとともに、機械可読可能な大規模テキストデータを形成する。

- ・画像データの拡充(15万点)
- ・OCRによるテキストデータ化(27万点)
- ・校正テキストデータの作成・公開(3,000点)

R6年度実績

デジタル画像データ (20,371点)

○近代自筆資料:島崎藤村『夜明け前』原稿

1929年から1935年にかけて『中央公論』誌上に連載された歴史長篇の原稿。 藤村自ら和本の形に仕立てた。

〇寺社所蔵の古典籍資料:世界遺産 春日大社

国文研の調査を踏まえ、歌書等資料を撮影予定(覚書締結・一部撮影開始)

○個人所蔵資料:射和文庫、石水博物館

○その他貴重資料

静嘉堂文庫美術館所蔵『平仲物語』

現存する唯一の伝本で、国の重要文化財に指定

校正テキストデータ作成

ワークフロー及びツール構築

126点(6,705 コマ)の校正済みテキストデータを作成。

OCRテキストデータ(23,933点)

国文研所蔵資料を対象に、OCRによるテキストデータを作成した。

米国・欧州等との連携拡充

北米に点在する日本古典籍の画像化を進め、且つ北米の研究者司書との関係を構築するための取組として、NCC(北米日本研究資料調整協議会)と連携し、「NIJL-NCC/CDDP Digitization Grant Program」を開始した。NCCを窓口とした画像作成実施機関が3機関採択され、うち1機関による画像作成を開始した。



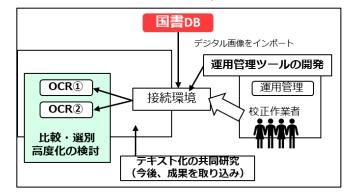
静嘉堂文庫美術館所蔵『平仲物語』 国の重要文化財に指定

> バリエーションある 文字データが抽出可 能なものを選定。

事例紹介

2025年度実績

校正テキストデータ



他ジャンルでも著名 な作品や教科書に掲 載のある作品を選定。 『紙漉重寶記』

「紙漉重寶記」
(国文学研究資料館所蔵)

出典: 国書データベース, https://doi.org/10.20730/200022057



『安政二乙卯年十月/江戸大地震出火塲明細記』 (国文学研究資料館所蔵)

出典: 国書データベース, https://doi.org/10.20730/200022082

2 異分野融合によるデータ駆動型研究の推進

異分野融合研究

萌芽研究(将来的な発展が期待される萌芽的な研究)、共同研究(成果の具体化が期待される研究)及びNW事業発展型 研究(NW事業において実施した共同研究に基づき、発展的に実施する研究)の区分を設定しており、2025年度から国文研DDH プロジェクトの重要事項を研究テーマとする重点課題研究 (テキストデータ関連) を新たに開始しました。

事例紹介

データインフラストラクチャの構築

共同研究

「AI技術を用いた大規模古典籍画像に対する新たな検索手法の研究」

国立情報学研究所及び法政大学との共同研究。事前に画像 にタグ情報をつけずにテキスト検索ができるCLIPを利用し、 自然言語による直感的な検索を可能とする試作システムを作 成。0.1秒に満たない時間で、「昔の京都で和服を着た若い女 性が楽しそうに微笑む絵」のような長文に対しても、妥当な 画像検索ができること等を確認。



「梅に鶯が鳴く絵」の検索例 (法政大学 藤田悟教授他)

『光琳画式』(神戸大学附属図書館所蔵)

人文系データ分析技術の開発

TEIに準拠した構造化 日本語テキストに適した構造化の方法を検討

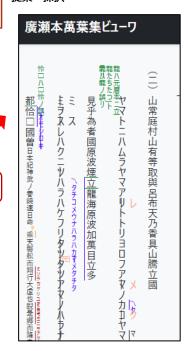
人文学資料を適切にデジタル化するための標準策定を 目的とした国際的共同プロジェクト。TEI コンソーシ アムにより運営。策定された標準はTEI ガイドライン Text Encoding Initiative として公開されている。

古典籍画像データから作成したテキストデータを扱いやす い形式に変換する共同研究を行うとともに、そのノウハウ も公開している。

多様な人々がデータ作成を担い、研究資源が増すことで、 異分野の研究者等にも人文情報が容易に利用できるように なる。

```
山常庭村山有等取與呂布天乃香具山騰立
  ▼<choice>
      <orig>國</orig>
                                           ブラウザ上で
      <reg>国</reg>
                                  日本語の縦書き・書き入れを再現
    </choice>
 </seg>
▼<seg type="rt" xml:id="note0100021" resp="#定家本
    ヤマトニハムラヤマア
 ▼<subst xml:id="shu0100032">
        <del type="ミセケチ" rend="red" resp="#廣瀬本
        <add place="right" rend="red" resp="#廣瀬本"
    トトリヨロフア
  ▼ <subst xml:id="shu0100033">
      <del type="ミセケチ" rend="red" resp="#廣瀬2
<add place="right" rend="red" resp="#廣瀬本"</pre>
```

※前身の歴史的典籍NW事業に おいて、TEIガイドライン「ルビ」を



(参考資料)



事例紹介

コンテンツ解析からの展開

共同研究研究代表者(宇宙物理学)が日本文学研究論文発表

道長家の〈花〉としての藤原妍子像	中	西	智	子 (1)
『保元物語』『平治物語』に見る古活字版刊行事業の一端 ············ 一第一種と第十一種を中小に一	阿	部	亮	太(3	9)
縁起と赤気 付・関連史料翻刻	片	畄	龍	峰(7	9)
1112 1111 11 11 11 11 11 11 11 11 11 11					_

連動

(国文学研究資料館紀要 文学研究篇, 第50号, 発行日 2024-03-22)

R6年度実績

残された古典籍からの アプローチ

1300年の蓄積

共同研究「天変地異の文理融合研究」

太陽極大期に伴い、日本各地でオーロ ラが見られたことから、最新のシチズ ンサイエンスで得られたオーロラ写真 の分析に、古典籍に残るオーロラの記 録を合わせて考察した論文を発表した。

シチズンサイエンス

(FExtended red aurora associated with super substorm igniting the October 10, 2024 magnetic storm as revealed by citizen science」Earth, Planets and Space 2025)



文献から太陽の活動履歴を知る 太陽から過去の災害の要因を知る 過去の文献から予知予防へ

マテリアル分析・解析

歴史的典籍のマテリアル分析

新たな分析手法の開発の可能性をも探る



古典籍の記録等

曲亭馬琴『近世物之本江戸作者部類。

世にこれを臭草紙といふ。この冊子は書皮に至るまで薄様の返魂紙 にて悪墨のにほひ有故に、臭草紙の名を負したり

「草双紙の本文料紙」は

「漉き返し(リサイクルペーパー)」

共同研究「マテリアル分析機器に基づく古典資料調査に関する総合的研究」

これまで館が蓄積したデジタル画像の内製モデ ルを活用し、他大学が所蔵する貴重書等を国文研 に搬入・燻蒸・搬出することと

して、実践女子大と 覚書を締結(2025年3月)。

R6年度実績

人間文化研究機構 国文学研究資料館 TR H K H G ма и кач 共同研究者(館外)が 微小部X線分析装置を利用する様子



R7年度実績