

大規模学術フロンティア促進事業（2024年度～2033年度）

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～

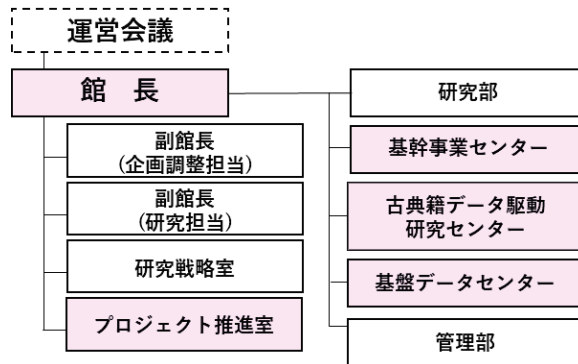
国文研DDHプロジェクト（略称）

科学技術・学術審議会学術分科会 人文学・社会科学特別委員会
令和6年8月22日



国文学研究資料館は、国内各地の日本文学とその関連資料を大規模に集積し、日本文学をはじめとするさまざまな分野の研究者の利用に供するとともに、それらに基づく先進的な共同研究を推進する日本文学の基盤的な総合研究機関です。創設以来培ってきた日本の古典籍に関する資料研究の蓄積を活かし、国内外の研究機関・研究者と連携し、日本の古典籍を豊かな知的資源として活用する、分野を横断した研究の創出に取り組みます。

組織図



建物外観



展示室

1951年5月
「文部省史料館」創設
(品川区)

1972年5月
国文学研究資料館 創設 (品川区)
「文部省史料館」を国文研の組織に組み入れ

2003年4月
総研大日本文学研究専攻 設置

2008年3月
移転(立川市)

2014年4月
大規模学術フロンティア促進
事業 開始
(歴史的典籍NW事業)

2024年4月
大規模学術フロンティア促進
事業 継続開始
(DDHプロジェクト)

2022年4月
古典籍データ駆動研究センター 設置

2004年4月
人間文化研究機構 設置

1950年代 1960年代 1970年代 1980年代 1990年代 2000年代 2010年代 2020年代

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（略称：歴史的典籍NW事業）

文部科学省の大規模学術フロンティア促進事業に**人文・社会系ではじめて採択**された歴史的典籍NW事業では、国文学研究資料館が中心となり、国内外の大学等の研究機関・図書館等と連携して**日本語の歴史的典籍30万点のデジタル画像化**を行い、諸分野の研究利用に資する**大規模データベースとしてWeb上で公開**し、その画像を用いた**国際的な共同研究のネットワークを構築**した。
(事業期間：2014年度～2023年度)。

これまでの取組と成果

研究手法の革新 「国書データベース」の構築・公開

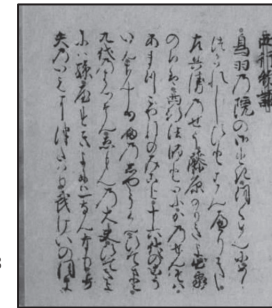


10年で30万点超



『西行物語』(静嘉堂文庫所蔵)
出典: 国書データベース
<https://doi.org/10.20730/100438358>
All-Rights-Reserved

静嘉堂文庫 重要文化財
現存最古『西行物語』書写本

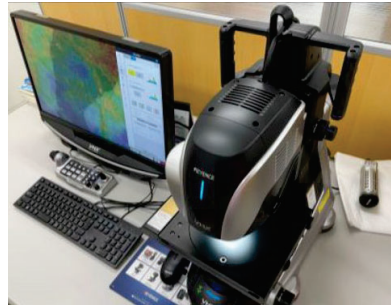


内製 — ノウハウの蓄積と公開



異分野共同研究等

共同研究の推進



くずし字データセット 古典に埋もれた「知」の発見

古典籍の科学的分析

	延べ(件)	区 分	(件)
成果発信数	1,813	公募型共同研究	5
うち著書	126	国文研主導共同研究	7
うち論文	684	拠点主導共同研究	4
うち口頭発表等	1,003	国際共同研究	7
	延べ(件)	異分野融合研究	17
共同研究参画者数	2,217	機構内連携共同研究	1
うち大学教員	755	研究開発系共同研究	19
うち若手研究者数	410	計	60

連携・利活用の拡充 ネットワーク等の構築

< 計136機関と覚書等を交わして連携 >

拠点機関

国外19、国内20

連携機関

①大学・高専51、②研究機関9、③学術団体6、④国公立図書館（博物館）・自治体（教育委員会）19、⑤その他（財団・企業・寺社・個人）12

国外（コレージュ・ド・フランス日本学高等研究所、大英図書館、ブリティッシュ・コロンビア大学、北京外国語大学等）
国内（北海道大、東北大、東大、名大、京大、阪大、九大、神戸大、広島大、慶應義塾大、早大、等）

①奈良先端大、東京海洋大、芸大、岡山大、実践女子大、学習院大、豊田工業高専 等
②極地研、情報研、データインテリジェンス共同利用基盤施設、機構内機関(歴博/国語研/日文研/民博)等
③日本近世文学会、中世文学会、中古文学会、日本学士院 等
④国立国会図書館、宮内庁書陵部、京都府立京都学・歴彩館、福井市、加賀市 等
⑤TOPPAN（株）、ポーラ文化研究所、清光山西厳寺、静嘉堂文庫 等

研究基盤の 維持・発展が不可欠・実験を実験で終わらせない更なる展開へ

大量の古典籍の電子化・テキスト化による一般への開放、及びデータの構造化・体系化と利活用

- ・ 社会の分断と疎隔
- ・ 古典との断絶
- ・ 現代の複雑な課題

なぜ今、古典の知が重要なのか。

千年以上にわたって引き継がれてきた
日本人の知識や知恵を、
現代にも通底する課題へのヒント
として利活用

人間社会の
課題解決

データ駆動による
課題解決型人文学

- ・ 古典の一般へ普及
- ・ 共感を生み出す
- ・ 古典の知恵を活かした課題の解決手法を開拓
- ・ 言語文化を継承

古典を開く
現代の課題から古典の知を発見する



データの
集積

データインフラ
ストラクチャの構築

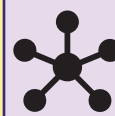
- 古典籍の15万点の追加画像データ
- 古典籍のテキストデータ
- メタデータ標準化 等



データの
解析

人文系データ分析
技術の開発

- AIの活用による大規模データの分析技術開発
- TEIによるデータの構造化 等



データの
利活用

コンテンツ解析
からの展開

- 社会や他分野に貢献可能なコンテンツの抽出
- 総合書物学や典籍防災学等の研究を推進 等



新たな
データ開拓

マテリアル
分析・解析

- マテリアルとしての書物の分析・解析技術を開発
- マテリアルデータのメタデータ化

データインフラストラクチャの構築

国書データベース

書誌検索

著作検索

著者検索

画像タグ検索

全文検索

試験運用

源氏物語画帖

光源氏



<https://kokusho.nijl.ac.jp/>

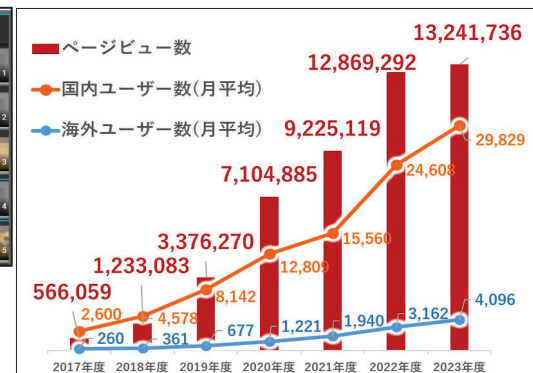


いつでも

どこでも
学問の進化

誰でも

拡 充



歴史的典籍とデジタル画像のメタデータ標準化

メタデータ流通ガイドライン

「メタデータ流通ガイドライン」(Metadata Distribution Guidelines)

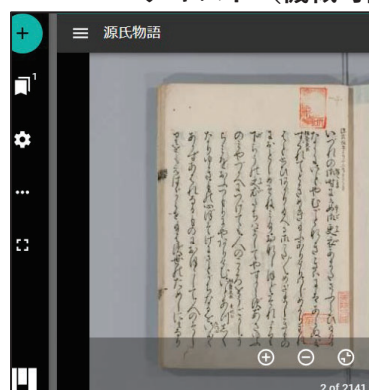
ガイドライン目次

- はじめに
- 共通編
- 研究データ編
- 古典籍編
- 別紙

国立国会図書館と当館が協力し、オープンアクセスリポジトリ推進協会(JPCOAR)等と連携して策定

メタデータ標準化

テキスト(機械可読)データの拡充



『源氏物語』(国文学研究資料館所蔵)
出典: 国書データベース,
<https://doi.org/10.20730/200010454>

AI

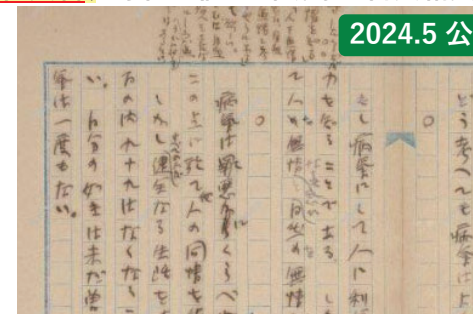
いつれの御時にか女御更衣あまたさかひ給けるなかにいとやむ二なききはあらぬがすてるときめき給ふありけりはじめよりわれはと思ひあがり給へる御たぐくめさましきものにおとしめそねみ給おなしほどそれより下らうの更衣たちはましてやすからずあさゆふのみやつかへにつけても人のこころをうごかしつらみをおふつもりにやありけむいとあづしくなりゆきもの心ほそけにさとがちなるをいよくあかずあはれるものにおほして人のそしりをもえはとからせ給はず世のためしにもなり

テキスト 27万点

OCRによる
テキストデータ化

近代資料、寺社・個人所蔵等の古典籍資料

2024.5 公開



武者小路の初期小説草稿。欄外に雑誌「白樺」の同人であり友人だった志賀直哉の書込有

『原稿：病氣』(調布市武者小路実篤記念館所蔵)
国書データベース<https://doi.org/10.20730/300047555>

All-Rights-Reserved

画像 15万点

近世(江戸)以前

近代(明治～終戦)

現代(戦後)

歴史的典籍NW事業

国文研DDHプロジェクト(+近代資料、寺社・個人所蔵等)

空白を
埋める

他機関デジタルライブラリー

古代から近代へとつながるシームレスな日本の書物のデジタル環境の構築を目指す

- ・ **テキストデータの可読性の向上**を目的として、TEIコンソーシアムの策定するガイドライン等に準拠したデータの構造化を推進
- ・ AIの活用により**大規模データから機図や画像などの情報を抽出**する非テキスト情報に関する技術開発を推進

TEIフォーマット作成及びツール開発：

- ・ 日本語環境での運用に適したツール開発
- ・ 既存ツールの日本語対応 等

メタデータ付与に関する汎用的仕組の開発及び国際標準に準拠したテキストの構造化：

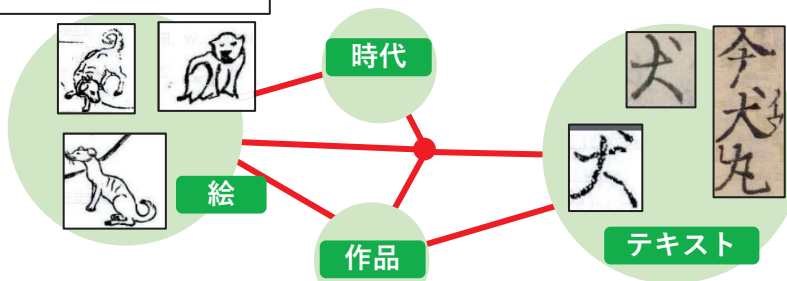
- ・ TEIに準拠した構造化テキスト作成
- ・ AI技術に基づく機械可読データ作成の効率（自動）化 **AI**
- ・ 構造化テキストの「日本語コーパス」等の外部データベースへの適用 等

画像検索・解析技術の精度向上と可視的把握技術の確立：

- ・ 非テキスト情報による検索技術の発展的な研究
- ・ 検索結果の可視的把握方法の検討・拡充

(例) テキストや絵などを自在に検索

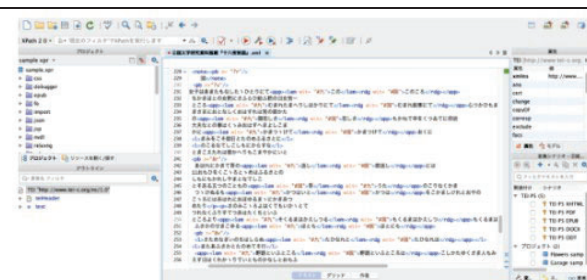
犬



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

TEI (Text Encoding Initiative) 規格による
日本語テキストの
機械可読化



```
176  かたはらにかきつく
177  <note><pb r="4r"/>
178  図</note>
179  <pb r="4v"/>
180  <!--とくめをくふるきまぐらのちりをたに</!-->
181  <!--われたちさらはたれかはらん</!-->
182  代々にかきをかきける<app><lem wit="#九">哥</lem><rdg wit="#国">うた</rdg></app>のさうしとも
183  のおくかきなとしてあたならぬかきり
184  を<app><lem wit="#九">えり</lem><rdg wit="#国">あり</rdg></app>したうめて<app><lem wit="#九">侍従</lem><rdg wit="#国">侍従</rdg></app>
185  てかきそへたる<app><lem wit="#九">哥</lem><rdg wit="#国">うた</rdg></app>
186  <!--とくめをくふるきまぐらのちりをたに</!-->
187  <!--われたちさらはたれかはらん</!-->
188  <!--とくめをくふるきまぐらのちりをたに</!-->
189  <!--とくめをくふるきまぐらのちりをたに</!-->
190  <!--とくめをくふるきまぐらのちりをたに</!-->
```

AIによる機械可読データの生成により
「個」の作品分析から
「束」としての解析へ

コンテンツ解析からの展開

多分野での歴史的典籍データの有効性

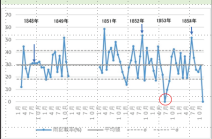
コンテンツ解析 (例)

茨城大学地球・地域環境共創機構共同セミナー (2024年3月)

古日記天候記録を用いた一九世紀半ばの山陰の気候の推定

おぼえ 大保恵日記

・1848年7月、1853年9月、1854年7月に洪水の記録
・1853年夏季には**干ばつ**の記録



日々の出来事

天候

地震

天文現象

災害の記録

比較 → 1849年 1854年に

暖冬傾向

降雪率記録が
同調

天候

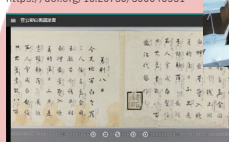
山陰の他の古日記

井上 貴之 (茨城大学理工学研究科博士前期課程)
野澤 恵 (茨城大学大学院理工学研究科(理学野)教授)

文献観光資源学

デジタルデータの収集と分析の成果
を地域に密着した観光資源として活用
／歴史的典籍データを観光資源へ
と転換し地域創生に寄与

『豊公御自撰集書』(東洋資料館所蔵)
出版: 国書データベース
<https://doi.org/10.20730/300046531>



国文研教員による
資料の現地調査(2020)

令和5年度廣瀬資料館セミナー
(2024.2 日田市複合文化施設)

中川文庫

(祐徳稲荷神社)

佐賀県鹿島市

第5回中川文庫セミナー
(2024.7 祐徳稲荷神社)

廣瀬資料館

大分県日田市

国書データベースにおいて資料を公開し、国文研基幹事業により共同研究を実施、地域に研究成果を還元している

典籍防災学

歴史的典籍データ等を
防災研究のエビデンス
データとして活用

藤原定家『明月記』の記述から、鎌倉時代にお
ける連発巨大磁気嵐の発生パターンを解明

アメリカ地球物理学連合「Space Weather」に掲載、新聞各紙
にも取り上げられた。

Space Weather

Historical space weather monitoring of prolonged aurora
activities in Japan and in China

Yoshio Kikuchi, Hirotaka Kido, Hisashi Higashiyama, Masafumi Tamahara, Akito Davis Kawanishi,
Masaki Moriguchi, Yohji Nishimura, Kazuhiko Yamamoto, Masaki Taki, ... See all authors

First published: 28 January 2017 | <https://doi.org/10.1002/2016SW001493> | CiteSpace 8

Abstract
Great magnetic storms are recorded as aurora sightings in historical documents. The
earliest known example of "prolonged" aurora sightings, with aurora persisted for hours

「Historical space weather monitoring of prolonged aurora activities in Japan and in China」
Space Weather <http://onlinelibrary.wiley.com/doi/10.1002/2016SW001493/abstract>

太陽風

太陽フレア

宇宙

地球への影響

- ・自然界への影響
 - ・災害との関わり
 - ・食物への影響
- たとえば
磁気嵐災害

日本最古の天文記録は『日本書紀』に
記された扇形オーロラ



日本最古の天文記録
として知られる『日
本書紀』推古二十八
年(620年)のくだり
に記された「赤気」
について、近年の古
典籍を用いたオーロ
ラ研究で解明されて
きた「扇形オーロ
ラ」と整合的である
ことを明らかにした。

国立極地研究所、総合研
究大学院大学との共同プ
レズリリース。

1000年の蓄積

残された古典籍からのアプローチ

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

総合書物学

人間文化研究機構人間文化研究創発セン
ター広領域連携型基幹研究プロジェクト

国文学研究資料館

古活字版の組成・版面パターンの情報工学的解析

国立歴史民俗博物館

延喜式のデジタル技術による汎用化

国立国語研究所

古辞書類に基づく語彙資源の拡張と
語彙・表記の史的変遷

総合研究大学院大学
「総合書物論」開講

国語研

『コンテツス・ムンチ』
全文テキストを公開※
2023.7

歴博

「デジタル延喜式」
正式公開2022.4

※ドイツのヘルフォーク・アウグスト図書館所蔵

社会や他分野に貢献可能なコンテンツの抽出

総合書物学や典籍防災学等の研究を推進

課題の具体相を見出す「問い」を生み出すデータとして、歴史的典籍データの積極的な活用を推進する

マテリアル分析・解析

マテリアル分析

マテリアルとしての書物の分析・解析技術の開発を行い、情報を抽出

デジタルマイクロ スコープ



墨付きの
繊維群・紙片

古典籍の記録等

曲亭馬琴『近世物之本江戸作者部類』

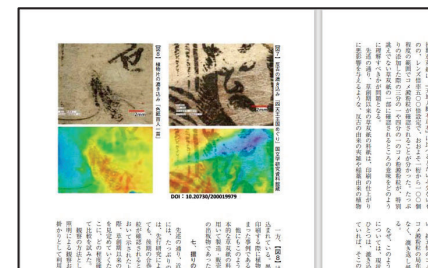
世にこれを臭草紙といふ。この冊子は書皮に至るまで薄様の返魂紙にて悪墨のほひ有故に、臭草紙の名を負したり

「草双紙の本文料紙」は
「漉き返し（リサイクルペーパー）」

科学的な証拠

分析結果

反古（不要になった紙）が使用されている



『近世文藝』117号（2023年1月）
査読付き論文掲載

「草双紙の本文料紙の紙質：高精細デジタル顕微鏡の観察結果を手掛かりに」松原 哲子

「先端研究支援ラボ（仮）」設置計画

古典籍データ駆動研究センターにおける
基盤的研究の成果を踏まえた発展的研究を推進

微小部X線分析装置

X線 回折X線を
検出器で測定



非破壊検査

ハンドヘルド 蛍光X線分析装置

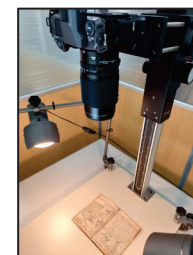


顔料等に含まれる
金属元素の分析

植物（原料）

毛髪など
人間由来の成分

赤外線カメラ



ブックスキャナ



<解析のイメージ>

- ・DNA解析で1000年間のヒトを知る
- ・和紙に含まれた毛髪から
当時の食生活を知る

マテリアル分析・解析技術の 高度化の推進

マテリアルデータを メタデータとして整備

モノに関する「研究者の経験値」を検証可能な新たなメタデータとして整備・蓄積・連携し、活用する

共同研究 第1期（計29件）

萌芽研究 4件

- 例) 「東アジアの歴史的な文字関連資料共有のためのデータインフラストラクチャの構築に関する準備研究」
- 漢字を中心とした東アジアの文字関連資料に関するデータセットを互いに連携
 - データセットの長期保存や低コストでの共有や利活用のためのWeb3技術や最新Web技術の適用可能性の検討

共同研究 8件

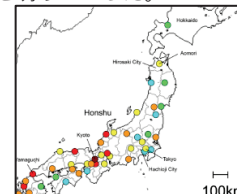
- 例) 「新古今和歌集の校訂本文データと現代語訳データの総合的研究」
- 新古今和歌集の校訂本文の作成、テキストデータ化。
 - 同作品の現代語訳データの作成、テキストデータへの紐づけ。

NW事業発展型研究 17件

- 例) 「古典籍と自然アーカイブの対比による宇宙気候研究」
- 古典籍に赤気や黒点等の記録が残る年代を中心に、確認されたイベントの太陽サイクルとの関連を解析
 - 古典籍の雷等の記録と太陽サイクルとの対比により、太陽活動の気候変動・気象現象への影響検討

歴史的典籍NW事業での成果

江戸時代の古典籍の記録等から、太陽の自転周期による過去300年にわたる日本の雷発生への影響を明らかにした。



現代の雷データ（1989～2014年）による27日周期の強度マップ（平成30年4月27日プレスリリース「太陽の自転周期が雷の発生に影響している―江戸時代の日記の分析で判明―」）

国際展開

日本古典籍研究国際コンソーシアム

- 日本古典籍を対象とした世界初のコンソーシアムとして設立。国文研が幹事機関。
- 日本国外42、日本国内41、合計83機関（2023年9月28日時点）。
- 代表委員会、分科会に加え、誰でも参加できる勉強会（10回（2023年度実績）「書誌学用語勉強会」）において日本古典籍を対象とした用語の多言語の検討を行っている。

書誌学用語勉強会・・・2024年度は、既に4回開催
海外向けの時間設定（9:00-10:00、21:00-22:00）



データベースに関連する国際コンソーシアムや学会への参加



IIFコンソーシアム



ジャパンリンクセンター (JaLC)



TEIコンソーシアム



日本デジタル・ヒューマニティーズ学会 (JADH)

- AAS 2023 Annual Conference（2023年3月）、NCC 2023 Open Meeting（2023年3月）、The 17th EAJRS Conference（2023年8月）、The 33rd EAJRS Conference（2023年9月）等の国際的アジア研究学会における研究および事業報告。

→ NCC（北米日本研究資料調整協議会）と連携し、
北米に所在する日本語の古典籍資料をデジタル化する取組を開始



Comprehensive Digitization and Discoverability Program
2024-2025 NIJL-NCC/CDDP Digitization Grant Program

若手・シニア研究者を交えた定期的なブリーフィング

若手研究者とシニア研究者、大学院生、技術職員、司書職員等を対象とし、積極的なアドバイスやコメントを得る場を設置、定期的に開催。

※2023年度開催実績

総合書物学の今後のプロジェクト内容の共有
古筆切資料のメタデータについて

「デジタル源氏物語」の共同開発について
DHを武器にする
大規模言語モデルと古典籍



人文学のための
データ構造化セミナー

人間文化研究機構 若手研究者派遣プログラム

平成28（2016）年度から機構本部実施の若手研究者海外派遣プログラムにより、毎年、若手研究者を海外研究機関に派遣（2020、2021年度は感染症拡大のため派遣中止）。

※2024年度派遣実績

派遣先	シェフィールド・ハラム大学(イギリス)
期間	2024.6.21～2024.7.2
研究課題名	テキストマイニングから読み解く物語の登場人物像 －「源氏物語」における六条御息所－

中高生向けSDGsイベント

2024.7.26実施



共同研究「古典資料の継承意識の涵養手法に関する教育的実践研究」の一環として、静岡聖光学院中学校・高等学校と共催で紙漉きと近世期和本の顕微鏡観察を体験するワークショップを開催。静岡聖光学院の生徒に加え、教育パートナーシップの維持・発展を目的とする「Experience Shizuoka Tour 2024」で滞在中のタイ・マレーシア・イタリアからの留学生も参加した。



小型顕微鏡や肉眼で
漉き返し紙の分析を行う参加者



キックオフシンポジウム（予定）

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～

K I C K O F F S Y M P O S I U M

A I × 人文学 — データ駆動による未来形成 —

- 日時：2024年12月1日（日）13：00～
- 場所：一橋講堂及びオンライン（YouTube配信）
- 主催：人間文化研究機構国文学研究資料館
- 後援：文部科学省



参考：2015年7月公開シンポジウム
（日本学術会議）

項目 (研究テーマ)	2024 (R6)	2025 (R7)	2026 (R8)	2027 (R9)	2028 (R10)	2029 (R11)	2030 (R12)	2031 (R13)	2032 (R14)	2033 (R15)	2033 (R16)
1. 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）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）											
	画像データの作成・公開										
	テキストデータ化										
	ワークフロー及び 作成ツールの開発			作成ツール改良、校正テキストデータの作成・公開							
2. 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以下の4つの研究計画により推進する。 ・データインフラストラクチャの構築 ・人文系データ分析技術の開発 ・コンテンツ解析 ・マテリアル分析・解析 多様な人文学分野の拠点である機構内各機関との協働により機構のスケールメリットを活かし、大規模データを活用した様々な課題意識に基づく国内外研究者との異分野融合研究を実施する。											
	共同研究第1期 (試行的研究)			共同研究第2期			共同研究第3期 (実装的研究)				
	対象資料の拡張に向けたデータ作成及び共同研究等の推進										
評価の実施時期	—	—	—	—	進捗評価	—	—	—	—	—	

期
末
評
価

期
末
評
価

日本語の歴史的典籍（古典籍）	慶應4年（1868年）以前に日本で出版・書写された書籍。
国書データベース	国文学研究資料館が令和5（2023）年3月より運用を開始した日本語の歴史的典籍を対象としたデータベース。目録データ（メタデータ）を中心とする「日本古典籍総合目録データベース」と、歴史的典籍NW事業により構築・運用していた画像データを中心とする「新日本古典籍総合データベース」との2つのデータベースを統合したデータを基盤として、新たにテキスト領域を設置している。
EAJRS（イー エー ジェイ アール エス） （「日本資料専門家欧州協会」）	「European Association of Japanese Resource Specialists」日本に関する情報や資料の開発と普及を促進・育成することを目的として設立された欧州の国際学会・団体。欧州の図書館員、研究者等が多く所属している。
AAS（エー エー エス） （「アジア研究協会」）	「The Association for Asian Studies」アメリカの学術団体でアジア研究者が多く所属している。
NCC（エヌ シー シー） （「北米日本研究資料調整協議会」）	「North American Coordinating Council on Japanese Library Resources」北米日本研究資料調整協議会(NCC)は、独立非営利教育団体として1991年に創設。北米の日本研究をより効果的に支援するために、個々の図書館では不可能あるいはそれを超えるレベルの日本研究資料の調整を中心に活動。多くの日本研究司書、大学教授、研究者等が協議会・委員会に参加している。
IIIFコンソーシアム（トリプルアイエフ コンソーシアム）	デジタル画像へのアクセスを標準化し相互運用性を実現しようとする国際的なコミュニティ活動。 https://iiif.io/community/consortium/
ジャパンリンクセンター(JaLC)	国際DOI財団に承認されたDOI登録機関であり、DOI登録サービスを提供する。 https://japanlinkcenter.org/top/
TEI（ティー イー アイ）	「Text Encoding Initiative」人文学資料を適切にデジタル化するための規格の標準策定を目的とした、国際的な共同プロジェクトであり、プロジェクト運営はTEI コンソーシアムによって行われている。策定された標準はTEI ガイドラインとして公開されている。 https://tei-c.org/
日本デジタル・ヒューマニティーズ学会	2011年設立。2013年からDH国際学会連合の構成団体。 https://www.jadh.org/