

特別号

2024年
3月発行

CONTENTS

「データ駆動による課題解決型人文学
の創成」事業の開始にあたって

国文学研究資料館長

渡部 泰明

①～③

数字でみる歴史的典籍NW事業

国文学研究資料館

古典籍共同研究事業センター長

山本 和明

④～⑥

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」の概要

プロジェクト推進室

設置準備タスクフォース

⑦

トピックス

⑧

ふみ

「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」ニューズレター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
古典籍共同研究事業センター「データ駆動による課題解決型人文学の創成」事業の
開始にあたって

国文学研究資料館長

渡部 泰明

わたなべ
やすあき

© えくてびあん

十年がかりで進めてきた、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」(以下、「歴史的典籍NW事業」)が令和五(二〇二三)年度で終了となる(最終年

度に収集した画像データの公開は次年度)。その後継計画に位置付けられる「データ駆動による課題解決型人文学の創成」事業(以下、「後継事業」と呼ぶ)も、事業移行評価を経て予算化され、いよいよこの令和六(二〇二四)年四月から本格的に始動する。「歴史的典籍NW事業」の実施に当たり、当館では日本古典籍研究国際コンソーシアムを設立し、これを当館が事務局を務めて推進しており、日本古典籍を対象とした書誌学の用語集(グロッサリ)を多言語で作り、オンラインで公開することを目指す書誌学用語勉強会や、コンソーシアム参加機関の専門職員・研究者・学生が情報共有する場として各種オンライン分科会を開催する等、国際的な取り組みを推進してきた。しかし何といっても、三〇万点を越える画像データの公開が、その国内外への波及効果を含め、研究の方法の革新につながる大きな成果と呼べるものであった。続く「後継事業」は次の五本の柱から成る。

- (1) データインフラストラクチャーの構築
- (2) 人文系データ分析技術の開発
- (3) コンテンツ解析からの展開
- (4) マテリアル分析・解析
- (5) データ駆動型人文学研究の展開

それぞれの説明は別の機会に譲り、画像データの作成―これ自体も継続し、次の段階に入る―がどのように新たに展開を迎えるかに絞りたい。収集公開した三〇万点のうち、機械可読が技術的に当面不可能なものを除き、そのほとんどをテキストデータとすることである。「後継事業」の主要な柱である。右に述べたように、「歴史的典籍NW事業」による画像データは、新しい研究基盤となり、研究に画期をもたらした。研究者コミュニティに裨益するところ大であった。次のフェイズは、日本文学及びその関連領域に限らない研究者への貢献が図られることになる。画像データを、日本古典を専門としない研究者へも、あるいは研究者ではない人にも利用できるように、人も機械も読めるテキストデータに転換すれば、古典のデータを活用しうる享受層は、一挙に拡大することが期待される。さらにテキストデータを構造化し、より高度な機械可読性を持たせる。

ではテキストデータからどのように研究は展開するのか。この場を借りて今現在思うところを語ってみたい。しばし夢に付き合っていたことを願う。

AIによるテキスト化と同時に、教科書や一般書で目にするような、一部について校訂本文の作成を行い、読みやすい本文を提供する予定である。続けて、主要古典作品の現代語訳に着手してみたい。校訂本文とその現代語訳が紐づけられたなら、古典に対する私たち現代人の向き合い方を一変させるだろう。古典の意義を変えらるだろう。社会生活のさまざまな局面から、逆に古典を想起することを可能にするのであるから。現代的な諸課題を起点に、古典本文へと遡行し、古典の知を発見することができる。発見こそ、あらゆ

る知の発動条件である。私たちは、この長い言語文化の歴史に、主體的に参画することができるようになる。

価値観は多様であった方が良いのは確かである。だがそれだけに、拠り所が確固としないと振り回されてしまう。多くの人々の支持と努力によって受け渡されてきた古典は、価値の源泉として、信頼するに足る存在だ。古典と現代語とが可逆的に連係することになれば、あらゆる分野の知と日本古典が、言語を介して接続可能な状態になる。巨視的に言えば、文化の歴史は古典再生の歴史といっても過言ではないから、再生した古典は次世代の文化の創成へとつながるだろう。人文知の新たな問いを生み出し、問いを探索する能力の高度化を主眼として、「日本文学の創成」ではなく「人文学の創成」を企て、「課題解決型」を冠した所以である。

私たちが今直面しているのは、分断と疎隔に絶えず脅かされ、浸食にさらされている社会である。この時代に人文学が果たすべき役割とは何か。もちろん答えは一つではない。が、少なくとも研究者には、「共感」の構造・機能を解明し、それが形成されるための基盤づくりに寄与することが求められている。

ここでいう「共感」とは、たんなる同情や思いやり（シンパシー）にとどまるものではない。立場・状況の違う相手の身になる能力（エンパシー）を指す。別個の生命体であることを前提に、その感覚や情動を我が事として受け止める器量のことだ。情においても理においても他者が理解できる能力であり、その身になれる能力である。その人の身になる。それこそ日本文化の根底にあるものだ。

もとより人はみな違うのだから、そっくり同じ心にはなれない。しかしその心境を想像し、その身になろうとすることはできる。むしろそう目指す営みのうちに、すこやかに「共感」は育つ。同情は人や集団に対して抱く一種類のものであることが多く、ゆえに一方への同情は、他方への疎隔を生みかねない。「共感」は複数のものが共存可能である。多様な存在に対して多様なままに応じうる。多様なものの共生は、現代の最重要の課題である。

ここで忘れてはいけないのは、「共感」とは発見によってもたらされる、という事実だ。「共感」は同じ気持ちであることを前提に形成されるものではない。探求し、想像して見出すものだ。「共感」とは発見して生み出すものだ。

「共感」を生み出すために、古典ほどふさわしいものはない。「共感」によって生まれ、「共感」に支えられて継承されてきた共通の文化遺産なのだから。といっても、ただちに胸を張るのは考えものだ。あまねく人々に浸透しているかといえば、いろいろ留保をつけざるをえない。教育という点では、現在初等・中等教育から取り上げられ、事実上ほぼ義務教育となった高等学校でも、必修科目『言語文化』には古典が含まれ、配当時間でいえばその半分以上を、古典を読むことに割くべきとされている（『高等学校学習指導要領 国語編』）。それだけの時間を充てても、皆が十分古典になじみ、興味を深めているとは言い難い。

思うに、高等学校の教育でいえば、現代語訳を重視しすぎている――文法偏重は論外である――のであろう。古典の現代語訳はかなり高度な技術を要する作業であり、これの学習を優先することによって、内容の魅力・興味深さを理解するまでに至りがたくなっている。現代語訳の設問は上級の応用問題と思った方がよい。過重な負担となれば、学習の足枷となる。訳は、基本的にどんどん示してしまっただけかまわれない、と思う。訳を参照しながら、自ら発見するよう促す。それが「共感」力を育てるカギとなる。

本事業計画は、古典教育の改善もしくは変革にも寄与する。正しく、わかりやすく、最新の成果に基づき本文を提供することに始まり、現代語訳も合わせて検索調査することができるようになれば、古典を全く知らない人間でも、アプローチすることが可能になるだろう。中等教育に限ったことではなく、関心を持つ人すべてに古典を開き、古典に「発見」的体験をもたらすのである。言語文化の継承を果たすことにつながるはずである。

「後継事業」開始に際して、まずは夢を語ってみた。共感していた

4つの研究領域からデータ駆動による課題解決型人文学へ

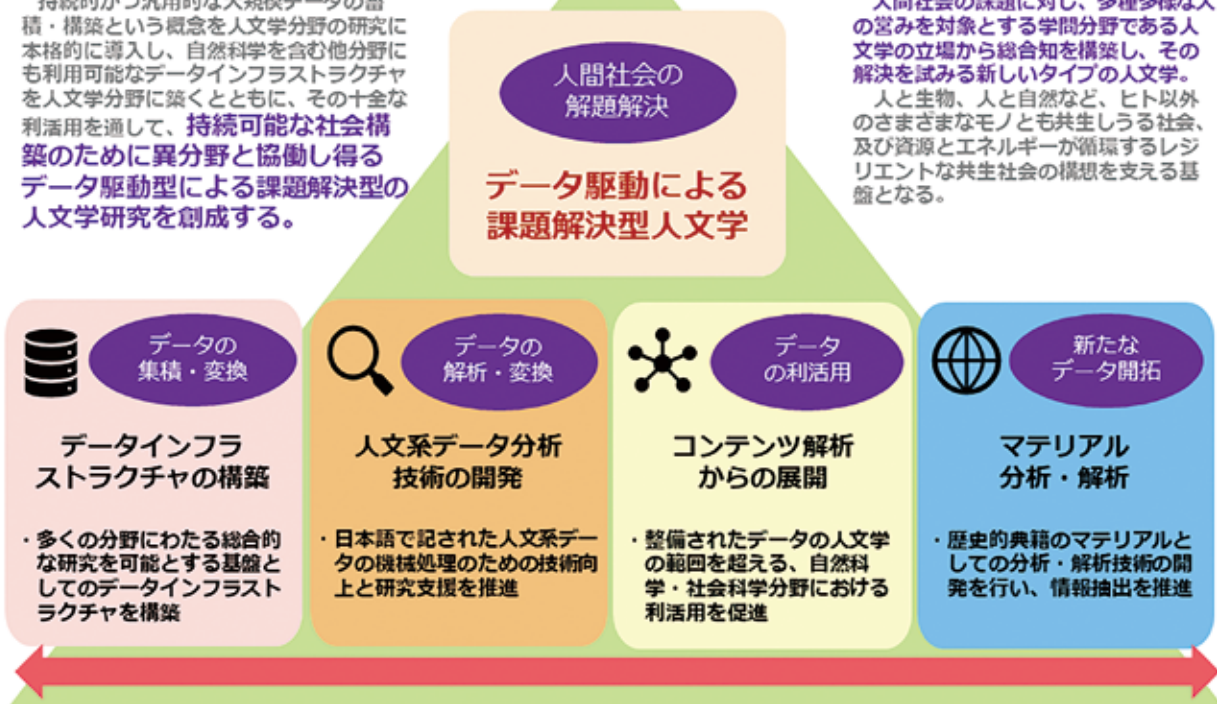
本計画の目的

持続的かつ汎用的な大規模データの蓄積・構築という概念を人文学分野の研究に本格的に導入し、自然科学を含む他分野にも利用可能なデータインフラストラクチャを人文学分野に築くとともに、その十全な活用を通して、**持続可能な社会構築のために異分野と協働し得るデータ駆動型による課題解決型の人文学研究を創成する。**

課題解決型人文学とは？

人間社会の課題に対し、多種多様な人の営みを対象とする学問分野である人文学の立場から総合知を構築し、その解決を試みる新しいタイプの人文学。
人と生物、人と自然など、ヒト以外のさまざまなモノとも共生しうる社会、及び資源とエネルギーが循環するレジリエントな共生社会の構想を支える基盤となる。

だけるなら幸いである。



計画の概要

数字でみる歴史的典籍NW事業

二〇二四年三月末をもって、当館が進めてきた大規模学術フロ
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」（通称歴史的典籍NW事業）は満期終了する。一年
の準備期間を経て、二〇一四年にスタートした一〇年プロジェクト
だ。本誌の発行元である古典籍共同研究事業センター（以下、セ
ンター）を中心に本事業が進められてきたが、一〇年をかけて達成
された事業成果については、当事業のモニタリング・評価をおこな
う日本語歴史的典籍ネットワーク委員会による『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―事業評価書―」（評価
書五四頁・事業評価書関連資料三五四頁、二〇二三年五月発行）が
詳細に纏めてくださっており、本誌二〇号でも紹介した通りであ
る。ページ数も多いがどうぞご覧いただきたい。ここで同じことを
繰り返すのは甚だ「芸」がないので、少し趣向を変えて、あまりお示
ししてこなかった「数字」から歴史的典籍NW事業を振り返り、少
し解説を加えてみたい。

*

*

二〇一三年の準備期間に、専従の教員が二名、事務（兼任、補佐員
含）が一七名の体制でセンターはスタートした。人文社会系ではじ
めての大規模学術フロンティア促進事業を、当館が担うにしても
甚だ小規模のスタートと言えるだろう。目指すのは、画像データを
三〇万点作成し、WEB上に掲載すること、異分野との共同研究を
進めること等々、明確であった。目標は定まっていた、それをど

国文学研究資料館 古典籍共同研究事業センター長

山本 和明
やまもと かずあき

のように進めれば良いのか。国文研に、新たに一セクションが出来
たとしても、全く経験のないことであり、霧のなかを彷徨うごとく
方針は拓けてはいなかった。大規模な施設整備計画とは異なるフ
ロンティア事業なのだから自分たちで考え、進めていかねばなら
なかったということだ。それが現在、教員六名（兼任含）、事務三四
名の体制にまで成長した。もちろん、館の事務やほかの教員の支援
も欠かせない。国文研を挙げての体制となった。センター事務増員
の大きな要因は、画像作成の量的拡大とそれに伴う書誌データの
入力作業、また画像撮影の内製化にともなう撮影班等の増員に尽
きる。満期を迎えるにあたり、三〇万点という目標に対し、日々立
ち向かってくれたセンター事務の人たち、作業にあたってくれた
補佐員たちあつての達成と感謝したい。国文研が元々持ち合わせ
ていたデジタル画像の点数は三万五千点ほど。それを一〇年で
三〇万点に増やすというのは並大抵のことではない。また、この事
業は毎年毎年の概算要求を経なくてはならない。次年度予算の内
示が出るのが一二月末。予算の見通しが立たなくとも、人を確保し
仕事をしてもらわなくてはならない。スタートした二〇一四年度
当時の予算は、一・五億円であった。人件費を含めての単年度予算
である。これに、国文研の経常経費からの捻出や人文機構のご支
援、外部資金（科研等）の獲得などを加えていくのだ（そのことすら
全く判っていないかった）。ちなみに事業計画では、当初所要経費は
総額八八億円と記されている（途中、実情にあわせ三九億五千万に

修正)。迂闊にも、年八億近くの資金を得てスタート出来ると思い込み、文字通り浮かれていたのだから、その落差は余りに大きい。今ではお笑い艸のハナシである。国に認められた事業計画である以上、是が非でも達成しなくてはならない。計画変更等が認められなければ、一〇年後あるいは中間評価時には、実施機関に対し「失敗」の烙印を押されることになる。一年目の画像作成数はマイクロコンピュータなどを含め一万八千点。前途多難のスタートだった。基盤整備としての画像作成は、今の時代となっても、AIで何とかなるものではなく、地道にデータ蓄積を進めていかざるを得ない。勿論、機械による自動ページめくりの可能性や撮影そのものの効率化などさまざまな工夫の余地は残されていたため、そうした検討や企業との共同研究も進めていった。和紙に書かれた(刷られた)古典籍という対象ゆえに、破損をしないことを第一義として、現在の撮影ノウハウを確定したのであった。内製(「自前で撮影していくということ」と業者撮影との併存という形で定着したのは二〇一八年頃のこと。単年度予算も二〇一九年度からは二億五千万円以上になり、それ以降、ほぼ年間三万八千点と言う目標数値をたて、現在に至る。なんとか三〇万点を達成しえたことで正直ほっとしている。当館のみで三〇万点の古典籍をもち、撮影していけるならまだしも、国文研の所蔵古典籍は二万点である。その他の二八万点余りは、所蔵機関との交渉・覚書等の締結・書誌データの作成などを依頼し、人的にも動いてもらう協力が不可欠である。撮影のご協力をいただいた各機関の方々からは、自館で作成してきた画像の提供などもいただけた。途中、新型コロナウイルスの蔓延により対応の難しい時期をなんとか乗り越えることが出来たのも、ひとえにそうした献身的なご協力あつてのものである。それなしには三〇万点の達

成などあり得なかっただろう。三〇万点の画像公開は、画像枚数に換算して約三二〇〇万枚に及ぶ。一〇年前、国立国会図書館や早稲田大学図書館などを除き、まだまだWEB上での画像公開数は少なかった。本事業で何とか世界に追いつけたのではないだろうか。この古典籍画像三〇万点という量的達成は、居ながらにしてある程度の基礎資料を入手できる環境の構築となったはずである。これは日本文学研究者のみならず、他分野の研究者、海外の研究者、一般市民へも開放されたものであり、今後一層利活用が推進されていくに違いない。ビューワ等の開発は日進月歩であり、その都度、衆目を集めるのであるが、何よりもコンテンツが充実していかなくては話にならない。その意味で研究基盤としてのデータ構築を進め、そのデータを維持していくことを国文研は今後も担っていくことになる。仮に業者撮影のみで一枚あたりの撮影経費一五〇円としたなら、それだけで四八億円になることを思えば、その規模感やセンターの創意工夫、地道な努力の程を判ってもらえたかと思う。

二〇二四年三月段階で、画像作成に協力いただいた機関数(個人蔵含)は一八二機関。現在のところ、当館の国書データベースTOPページ右上の「…」をクリックすることでそのあらましを確認できる。

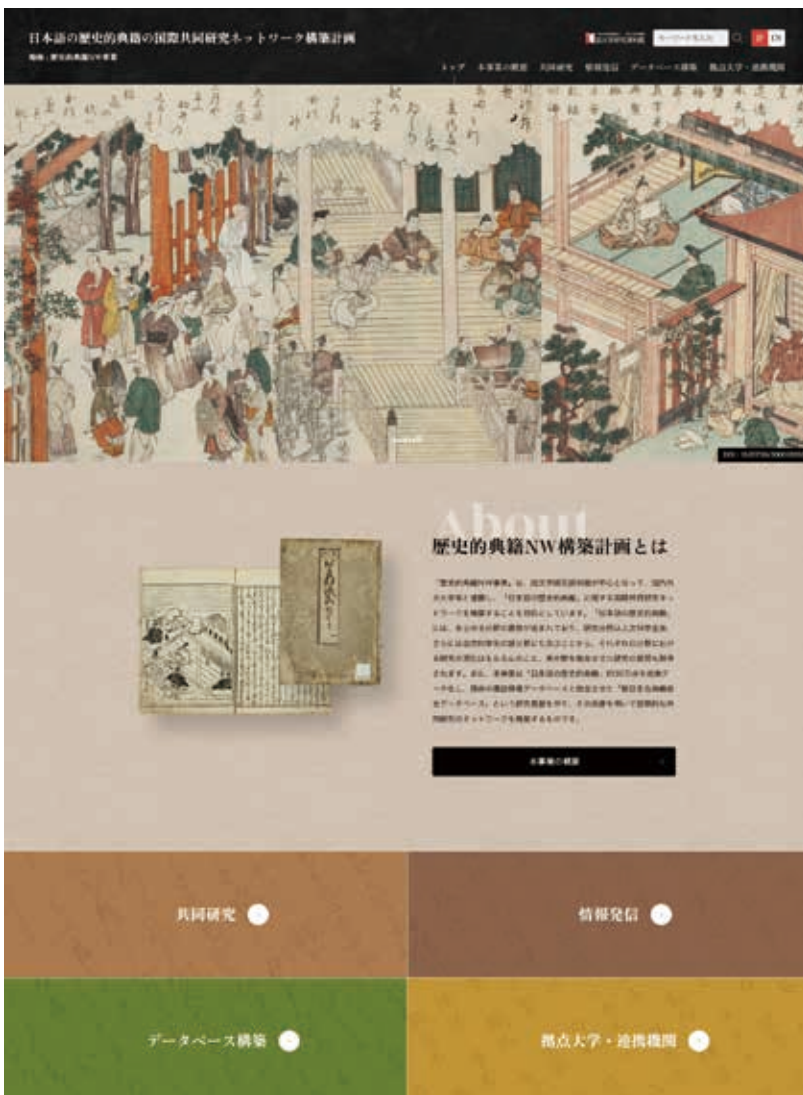
異分野融合研究を含め、共同研究のことも少し触れておこう。一〇年間で実施した共同研究数は六〇件、だいたい期間三年を目処にして実施された。くずし字認識に関する研究(TOPPAN、CODH、はこだて未来大学等)や、オーロラなど、典籍防災の研究(国立極地研究所、武蔵野美術大、茨城大学等)は様々に反響を呼んだ。他分野・他機関との共同研究、研究開発系共同研究など、一〇年

間で延べ二千二百人近くの研究者に参画していただいた。他分野との共同研究の難しさは、各々の研究分野の「常識」を知ることから始めなくてはならない点にある。相手にやってもらうのではなく、一緒に進めるためにはどうしたら良いか。未だに試行錯誤である。たとえば、プレ・プリントなどを活用し、いち早く成果を示すことを重視する分野と、冊子体にすることを重視し、年に二回程度の刊行誌に掲載を目指す人文系分野との相違点などは今後真っ先に埋めてかからねばならないと思う。

こうした研究成果や画像作成についての広報も積極的に展開していった。大学共同利用機関に位置する当館の存在などは、日本文学研究者やその周辺領域の研究者には認知されていても、一般市民の知るところではないと、もっと自覚を持たねばならないことだ。この事業期間にメディアリリース（新聞社などの報道機関に向けての広報）をした回数は四七回。メディアの人が興味をもってもらえるかという観点で「見だし」に留意するよう徹底したのだ。得てして広報という場合、情報を提供することにのみ注視しがちだが、そのリリースを受けて、取材を受け、新聞等の紙面に掲載してもらうまでを考えるべきだろう。当方で確認しているメディア掲載数は約三〇〇回（同一情報から様々なメディアに掲載された場合、各々でカウントしている）。研究や事業について如何に市民に還元しているか、である。レスポンスあつての情報発信を心がけたことで、この事業について多くの方に知っていただけたし、本事業については、

今後も継続し、当館HP上にアーカイブ化を進めることで、継続してご覧いただける予定である。

この歴史的典籍NW事業は終了するが、引き続き大規模学術フロンティア促進事業として、一〇年計画の「データ駆動による課題解決型人文学の創成」がこの四月からスタートする。館内組織も再編成され、センターも「基盤データセンター」と名称を変え、一五万点の画像、AI等活用してのテキスト化を担っていく。本誌「ふみ」も館内担当部署は変わるが、引き続き刊行される。今後の国文研の活動に、注視いただくとともに、ご支援を賜れば幸いである。



歴史的典籍NW事業アーカイブ画面

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」の概要

プロジェクト推進室設置準備タスクフォース

国文学研究資料館は、令和六年度から、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」(以下、「後継事業」)を実施していきます。

この後継事業は、文部科学省の大規模学術フロンティア促進事業として、平成二十六年度から令和五年度までの十カ年で実施した「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(以下、歴史的典籍NW事業)」の成果をさらに発展させるプロジェクトであり、令和六年度から令和十五年度までの十年計画となっています。

歴史的典籍NW事業から後継事業への移行に当たっては、科学技術・学術審議会 学術分科会 研究環境基盤部会 学術研究の大型プロジェクトに関する作業部会による事業移行評価を受けた後、令和六年度概算要求を経て、事業開始の運びとなりました。

実施体制については、国文学研究資料館長をリーダーとする「プロジェクト推進室」において事業全体のマネジメントを行い、その下に「古典籍データ駆動研究センター」、「基盤データセンター」、「基幹事業センター」の

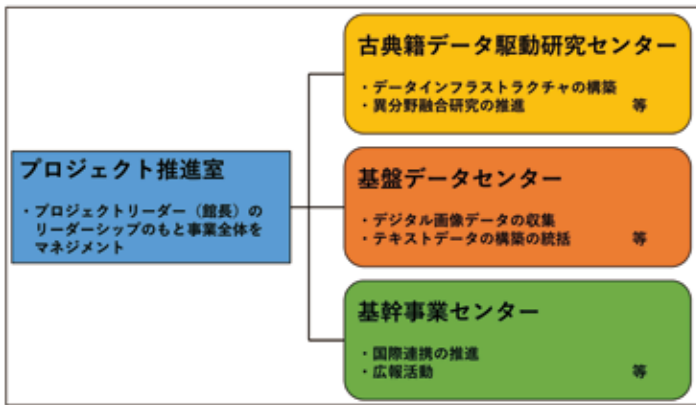


図1 組織図

三センターを設置(図1)し、三センターの緊密な連携により、後継事業を推進します。

二つの大きな研究テーマとして、「研究基盤データベースのデータ拡充・高度化」と「異分野融合によるデータ駆動型研究の推進」を掲げています。「研究基盤データベースのデータ拡充・高度化」では、歴史的典籍NW事業で構築した国書データベース(図2)を基盤として、古典籍十五万点の画像データの追加、画像データのテキストデータ化、データベースの機能追加などに取り組みます。また、「異分野融合によるデータ駆動型研究の推進」では、マテリアルとしての書物からの情報抽出を目指す研究などをはじめとして、データインフラの活用による異分野融合研究の進展等を図ります。

これらにより、他分野と協働し得る課題解決型の人文学研究の創成を目指して参ります。引き続き、関係の皆様のご支援とご協力をお願いいたします。



図2 国書データベースTOP
(<https://kokusho.nijl.ac.jp/>)

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 10年のあゆみ

事業開始（2014年4月）



市民参加型ワークショップ
「古典」オーロラハンター



「新日本古典籍総合データベース」公開（2017年10月）
画像公開点数 10万点を突破（2020年2月）

古典籍共同研究事業センターの事務職員



事務室



内製化室

「日本古典籍研究国際コンソーシアム」設立（2020年11月）

画像公開点数 20万点を突破（2022年10月）

「国書データベース」公開
（2023年3月）



国書データベース

画像作成点数30万点を達成（2024年3月）

（※2025年3月までに公開作業を完了予定）

2024年度からは本事業の後継事業として「データ駆動による課題解決型人文学の創成」を推進

事業・研究成果の一例



コラボレーション企画



ROIS-DS CODH開発
AIくずし字認識アプリ「みを(miwo)」



文献観光資源学
「津軽デジタル風土記の構築」プロジェクト
『津軽デジタル風土記資料集』

ふみは、後継スター
ト号として次回発行
予定です

■表題の背景色は二藍（ふたあい）です。二藍（ふたあい）とは、平安時代に愛好された紫系統の色ですが、着用者の年齢などによって色調が異なります。『源氏物語』では、光源氏が息子の夕霧に対し「二藍もいいが、余り赤みが強いものは見くらげられるだろう」と忠告する場面があります。

■本誌「ふみ」各頁の背景は当資料館蔵の『方丈記』（本阿弥光悦流の書体を模刻した嵯峨本）を利用しています。

■表題「ふみ」の書体は、石川島造船所（現IH）創業者の平野富二が明治十二年六月に刊行し当館所蔵の「BOOK OF SPECIMENS」（活版印刷見本帳）を利用しています。

ふみ

「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」ニューズレター
特別号

（発行日）

令和6（2024）年3月15日

（編集・発行）

国文学研究資料館

古典籍共同研究事業センター

〒190-0014

東京都立川市緑町十一三

TEL 050-5533-2988

FAX 042-526-8883

<http://www.nijl.ac.jp/pages/cijproject/>