

歴史的典籍NW事業の現在と未来

画像集積からデータ駆動へ

宮本祐規子 & 海野圭介

国文学研究資料館 古典籍共同研究事業センター特任准教授

国文学研究資料館・研究主幹／後継計画準備室長

第6回日本語の歴史的典籍国際研究集会
The 6th International Conference on Pre-modern Japanese Texts

2020年11月7日（土） オンライン開催

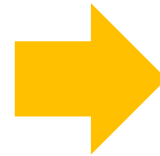
<https://www.nijl.ac.jp/pages/cijproject/sympo2020.html>

主催：国文学研究資料館

学術研究の大型プロジェクト

現在推進している大規模学術フロンティア事業
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
平成26(2014)年度～令和5(2023)年度

後継プロジェクトとして立案
データ駆動による
課題解決型人文学の創成



『学術研究の大型プロジェクトの推進に関する基本構想 ロードマップの策定
ーロードマップ2020ー』

文部科学省より資料公開（令和2年9月24日）
https://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/toushin/1388523_00001.htm



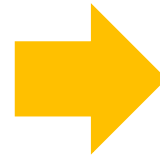
学術研究の大型プロジェクト

後継プロジェクトとして立案

データ駆動による

課題解決型人文学の創成

『学術研究の大型プロジェクトの推進に関する基本構想 ロードマップの策定
ーロードマップ2020ー』



文部科学省より資料公開（令和2年9月24日）
https://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/toushin/1388523_00001.htm

個々の研究者による個別の研究蓄積と経験に基づき進められることが主流であり、多分野間で共有され、将来にわたって継承される研究データの蓄積という概念の薄かった人文学分野の研究を、自然科学を含む他の分野にも共有し得るデータ駆動型に再構築し、そうした個々の研究蓄積を総合することで現代社会のかかえる大きな課題に対応するためのツールとなる、さまざまなデータを汲み出すことのできるデータインフラストラクチャーを人文学分野に築き、他分野とともに参画できる課題解決型の人文学研究という新たな知の創成を目指す。

画像データを発信する



新日本古典籍総合
データベース
<https://kotenseki.nijl.ac.jp>



情報の発信

30万点の古典籍
の全文画像データ



画像データを検索する



新日本古典籍総合
データベース
<https://kotenseki.nijl.ac.jp>

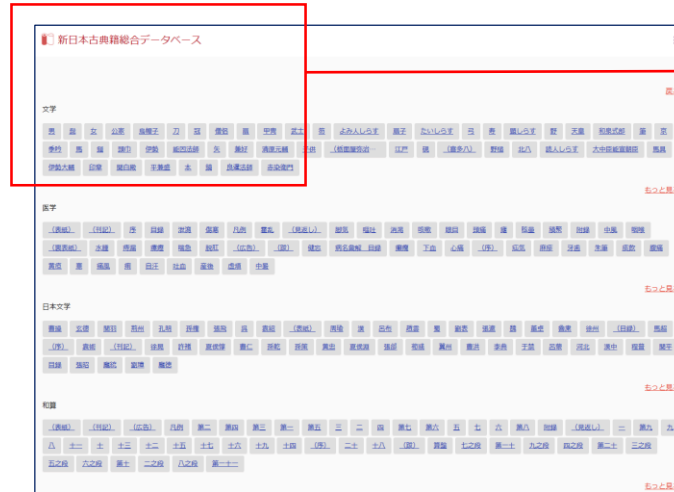


情報の検索

30万点の古典籍
の全文画像データ
の中から情報を検
索して取り出す



タグを利用して画像を検索する



- * 人間の目と手による情報の付加
- * 人間の目と手による情報の検索

30万点の歴史的典籍の情報量とは？

- 30万点の歴史的典籍が、各100丁（200ページ）だったら？
- 30万点×100コマのデジタル画像データ = 3000万コマの画像
- 1コマを10秒で目を通すとして…

1分で6コマ

1時間で360コマ

1日で8640コマ

1月で259,200コマ

1年で3,153,600コマ



3000万コマの画像データを
読むには、約9年半かかる
とうてい1人の人間の読むこ
との出来る範囲を超えている

データを検索し活用する

?

何を知りたいの？



新日本古典籍総合
データベース
<https://kotenseki.nijl.ac.jp>

コンピュー
タ処理が可
能なデー
タの構
造化



データ処理
& 解析技術
の革新



現代社会の包摂する多様な
課題への対応と解決を目的
としたデータ駆動による課
題解決型の人文学の構築

データの拡充
コンテンツの追加

データを 有効活用する



新日本古典籍総合
データベース
<https://kotenseki.nijl.ac.jp>



人文学の蓄積を現代社会の課題解決へ

- データの蓄積からデータ駆動へ → 画像から機械可読データへ
* 蓄積されたデータは人文学の重要なインフラでありツール
- 人の手と目からコンピュータへ → 分散型から集約型へ
* 個別の研究者に紐付けられた蓄積を将来へわたって共有できるデータに
- 自然科学・社会科学と共有できるデータ構造 → 多分野融合
* 人間社会をめぐる課題の複雑化、広範囲化に対応するには分野横断で対応



1000年におよぶ日本の記憶と経験を世界にひらく

セクション③
小風尚樹先生
チームの報告

典籍に基づく課題解決型人文学で持続可能な社会の実現へ

4つのパースペクティヴから 現代社会の課題解決へ

セクション①
佐藤悟先生の
報告

データ駆動型人文学の創成

データ処理
の技術の革新
ガイドラインの策定
etc.

人文系データ分析

- 機械可読化

画像検索



NIJL
人文学データ駆動
研究センター設置

マテリアル分析
データから資源発見

DNA解析



物質として
の書物に含
まれるデー
タの解析

課題解決型
融合研究の
展開 etc.

コンテンツ解析
異分野融合研究の
国際展開

データ
インフラストラクチャー
IDによるデータ管理

データ拡大・変換

- 通時的データ収集
- AIによる画質向上

データの拡
充と変換技
術の革新
データイン
フラの開発

準備状況

基盤整備：30万点の全冊画像公開
研究：機関間で異分野融合研究

セクション②
田村隆先生
チームの報告