



大学共同利用機関法人

人間文化研究機構

資料 2 - 4

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（第7回）  
令和3年6月28日

# 人文情報学（DH）の可能性

大学共同利用機関法人  
人間文化研究機構 国立歴史民俗博物館  
後藤 真

# 現状把握 データの蓄積

- 人文資料についてのデータ蓄積は比較的進捗がある  
→ 「デジタルアーカイブ」
- 一方、研究（プロセス）データについては進展があまりない
- DH研究もまさに「これから」



# 人文情報学への試みの例（大型機関）

1. 国文学研究資料館・新日本古典籍総合データベース
2. 情報システム研究機構・ROIS-DS人文学オープンデータ共同利用センター
3. 東京大学史料編纂所・データベース群
4. SAT（大正新脩大蔵経データベース）
5. 国立歴史民俗博物館 khirin
6. 「みんなで翻刻」
7. 国立国語研究所・一連のコーパスセット
8. HNG（漢字字体規範史データセット）およびCHISE（京都大学）
9. 立命館大学・アートリサーチセンター

国文学研究資料館では、国際的な共同研究ネットワークの構築に向けた歴史的典籍NW事業に、平成26年度から取り組んでいます。令和2年3月末現在で、歴史的典籍14万8千点、コマ数では1850万コマを撮影しました。今後、毎年3万8千点の撮影を進め、30万点の画像公開を目指します。

現在、当事業センターは在宅勤務を推奨しており、電話での対応が難しくなっております。可能な限りメール ([cijinfo@nijl.ac.jp](mailto:cijinfo@nijl.ac.jp)) でのお問い合わせをお願いいたします。

## お知らせ

2021年3月30日 英文オンライン・ジャーナル Studies in Japanese Literature and Culture Volume 4: INTERACTION OF KNOWLEDGE を刊行いたしました。  
<https://www.nijl.ac.jp/pages/cijproject/sjlc.html>

2021年3月22日 当館と、米国ワシントンDCの国立アジア美術館 (スミソニアン協会フリーア美術館、アーサー・M・サックラー・ギャラリー) は、フリーア美術館所蔵「ブルヴェラー・コレクション」の高精細度画像を、新日本古典籍総合データベースに提供することで合意しました。2021年3月26日に最初に提供された12点を新日本古典籍総合データベースで先行公開いたします。  
[国文研プレスリリース](#)

2021年3月17日 千葉大学附属図書館 (千葉県千葉市) と「日本語の歴史的典籍に関する国際共同研究ネットワーク構築」を推進する覚書を2021年2月22日に締結しました。千葉大学附属図書館が所蔵する古典籍を新日本古典籍総合データベースで公開していきます。

2021年3月16日 国田学園女子大学 (兵庫県尼崎市) と「日本語の歴史的典籍に関する国際共同研究ネットワーク構築」を推進する覚書を2021年2月18日に締結しました。国田学園女子大学が所蔵する古典籍を新日本古典籍総合データベースで公開していきます。

[すべてのお知らせ](#)

## シンポジウム・イベント

**新日本古典籍**  
総合データベース

**新日本古典籍**  
総合データベース

パンフレット

古典籍のデジタル化  
撮影マニュアル

「くずし字データセット」  
データ作成基本仕様

日本語の歴史的典籍  
国際研究学会

ニュースレター ふみ 最新号

# 日本文学を中心とした「典籍」とよばれる歴史的書籍の総合デジタルネットワーク

## 大規模学術フロンティア事業

## CODHへの展開によるデータ解析へも

## 異分野融合事例としては日本で見たオーロラ(赤気)の分析例など

藤原定家  
日記の「赤気」はオーロラ 極地研などが解析

社会 環境・科学 探検 科学・テクノロジー

毎日新聞 | 2017/3/22 13:40(最終更新 3/22 17:40) | 有料記事 675文字



平安・鎌倉時代の歌人、藤原定家(1162~1241年)が日記「明月記」に書き残した「赤気(せつき)」という現象は、太陽の異常な活発化によって京都の夜空に連続して現れたオーロラだった可能性が高いと、国立極地研究所や国文学研究資料館などのチームが米地球物理学連合の学術誌に発表した。連続したオーロラの観測記録としては国内最古という。

明月記には、1204年2~3月にかけて、京都の北から北東の夜空に赤気が連続して現れ、定家は「山の向こうに起きた火事のように、重ね重ね恐ろしい」と書き残している。

# 後継事業

どんなプロジェクトなのか

データ駆動による  
課題解決型  
人文学の創成

国文学研究資料館の  
チャレンジ!

prologue  
世界どこからでも  
日本語の歴史的典籍を  
見ることができるようになった  
その先に

Section.01  
テキスト  
機械可読型にデータを整備する



## 【計画の概要】

# 1000年にわたる日本の記憶と経験を世界に開く

気候変動・地球環境史の記録、危機社会再生の経験、  
異文化接触・多文化共生の記憶、心の問題対処の歴史…

計画名：データ駆動による課題解決型人文学の創成



欧米と並び、連携するデータインフラ

## データの構造化

データ駆動で異分野・海外と共創

データ集積からデータ駆動へ

異分野融合で

次のステップへ



NIJL-NWプロジェクト：30万点の古典籍全冊

デジタル画像配信（オープンデータ）



国文研の  
ミッション

50年におよぶ古典籍の書誌情報&  
所在情報収集から情報基盤を構築

# 【計画の学術的意義】



## データ駆動型から 新分野の創成へ〈挑戦〉

典籍人類学  
典籍防災学  
文献観光資源学

分散型データ集積・運用に  
基づく人文学研究から

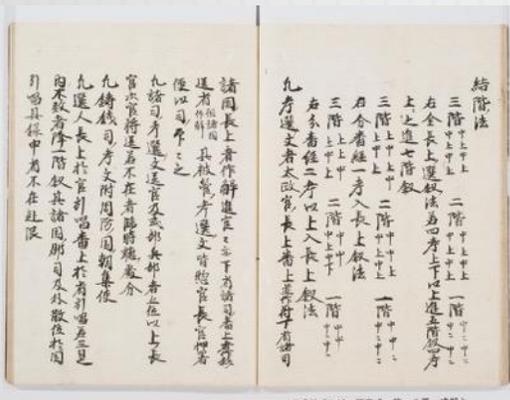
さまざまな形で得られたデータをもとに  
活用可能なモデルを見出す

## 「データ駆動型」の科学へ

データ利活用を通して異分野と協働し得る学問  
《課題解決型の人文学研究》へ国際展開

# 人間文化研究機構で進める「古代の百科全書・延喜式」のデータ構築例

## データからのVisualization



H-743-74-18 延喜式 第一八册 式部上

資料画像を含む複数画像をもとにテキスト構築 (TEI)

- 原文
- 現代語訳
- 英訳

```
<div type="条" n="24.26" ana="主計上">
<head ana="下総国"/>
<p>
<placeName xml:id="下総国">下総国
</placeName><行程、上卅日、下十五日、>
調、<measure xml:id="調_下総">
commodity="絁" quantity="200" unit="疋">絁二百疋</measure>、紺布六十端、
縹布卅端、黄布卅端、自余輸布、庸、輸布、
中男作物、麻四百斤、紙、熟麻、紅花、
</p>
</div>
<div type="条" n="24.27" ana="主計上">
<head ana="下総国"/>
<p>
<placeName xml:id="下総国">下総国
</placeName><行程、上卅日、下十五日、>
調、<measure xml:id="調_下総">
commodity="絁" quantity="200" unit="疋">絁二百疋</measure>、紺布六十端、
縹布卅端、黄布卅端、自余輸布、庸、輸布、
中男作物、麻四百斤、紙、熟麻、紅花、
</p>
</div>
```



Table for Checking the Accounts Balances

品名	延喜式	現代語訳
絁	200疋	200疋(約200kg)
紺布	60端	60端(約1200m)
縹布	30端	30端(約600m)
黄布	30端	30端(約600m)
自余輸布	40斤	40斤(約200kg)
庸	30端	30端(約600m)
輸布	30端	30端(約600m)
中男作物	400斤	400斤(約2000kg)
紙	400斤	400斤(約2000kg)
熟麻	400斤	400斤(約2000kg)
紅花	400斤	400斤(約2000kg)

Tables for Checking the Original Contexts

品名	延喜式	現代語訳
絁	200疋	200疋(約200kg)
紺布	60端	60端(約1200m)
縹布	30端	30端(約600m)
黄布	30端	30端(約600m)
自余輸布	40斤	40斤(約200kg)
庸	30端	30端(約600m)
輸布	30端	30端(約600m)
中男作物	400斤	400斤(約2000kg)
紙	400斤	400斤(約2000kg)
熟麻	400斤	400斤(約2000kg)
紅花	400斤	400斤(約2000kg)

画像データ (IIIF)



テキストを「読む」ことによる人文学的成果

- ここからさらなる研究へ (あくまでも可能性の例)
- 古代における食料分析 → 環境変動
  - 古代国家の経済構造解明
  - 合金技術分析
  - レシピ?



# DH研究のあり方とは何か（私見）

- 単に一つ一つの資料を読むことに加えて、新たな可能性を開く
  - Visualization, テキスト計量解析, データの組み合わせ, 研究プロセス可視化など
- 最後の本質を全て変える必要もない →DHは人文学
  - 人間が人間のことを考え、人間社会をよくするという人文学の行為は変えようがない
- 一方で、デジタルデータ化の新たな流れへと対応することで、より広がりがある研究（総合知への貢献）や、研究速度の向上へも期待シーズを活かしつつ、研究ニーズに合わせた展開を

# DH研究の推進のために必要なポイント

パッケージとしての  
推進が必要

## プラットフォーム・高度化

- 適正で安定的なプラットフォーム
  - オープンサイエンスのために
- 相互運用性
  - 共通フォーマットの利用・東アジアからの提案
- 高度な情報学的アウトプット
  - 分析技術の高度化

## 人文学の課題解決として

- 研究に基づく信頼できるデータ
  - 人文研究にもとづくデータ構築  
→元データが誤りだと結果も誤り
- 適切な分析技法
  - 情報学と人文学の適切なコラボレーション
- データ人文学的課題解決研究
  - データ構築も含めた分野と国を超えた人文研究アウトプット