

データ駆動による課題解決型人文学の創成

① ビジョンの概要

古典籍には未知の社会問題や予想外の危機に対処するための〈知〉が豊富に含まれている。近年デジタル化の急速な進展に伴い、これまでの人文学の方法では分析しきれない膨大なデータが文献情報資源として蓄積されつつある。古典籍も、現代に活かすデータという観点で捉え直し人工知能（AI）の活用と自然科学との協働による新たなデータ駆動型研究を創成することで、先人達の英知を〈総合知〉として現代社会に生かすことができる。

② ビジョンの内容

古典籍に含まれる〈知〉には、大きく分けて（１）事実、（２）営為、（３）思想（心情）の三種類があると考えられる。（１）（２）（３）は密接に関係していて、独立しているわけではないが、研究の精緻化に伴って、分野間相互の関係性が薄くなってしまおうという側面があった。また、良く知られた著名な古典だけを長年分析し続けるといった面もあり、研究が〈総合〉とは逆の隘路にはまる弊害が多々あることも否めない。

さらには、古典籍のデジタル化が進展し、デジタル画像やその電子テキストが膨大に蓄積（古典籍ビッグデータ）されるようになってきた。しかし、そのスピードがあまりに急速であったため、古典籍ビッグデータをどのように扱うかという方法が開拓されないままになっており、喫緊の課題となっている。

古典籍に基づいた人文学研究は、文理融合の研究を推進することで、その可能性を拡大していく余地が十分に残されていると考えることができる。また、媒体としての古典籍は、物質としての特性を持つ。世界的には稀なことではあるが、日本には多くの古典籍が残存している。これまでは文学や歴史分野においてテキストのみが研究対象とされ、物質としての面は捨象されてきた。しかし、昨今、料紙やそこに漉き込まれた毛髪の科学分析などをとおして、考古資料と同様の価値があるものとして検討が開始されつつあり、古典籍は研究資源として文理双方から注目されるところとなっている。

近時、日本古典籍のデータベースは大変充実してきた。パンデミックを経験して、研究者も個別の研究にそれらを利活用することは普通になってきている。しかし、データの全貌を分析し、研究することはまだ緒についたばかりである。厚みのある〈知〉をデータ基盤に移して蓄積し、人文学、自然科学の各分野と連携することは、文理異分野融合によるデータ駆動型研究にふさわしい、大きな可能性を持っていることは間違いない。このような研究を創成することで、人間や社会の統合的理解と課題解決に資する〈総合知〉に至る道を切り開くことができると確信する。

③ 学術研究構想の名称

データ駆動による課題解決型人文学の創成

④ 学術研究構想の概要

本計画は、経験則が重視され、持続的かつ汎用的な大規模データの構築・蓄積という概念の薄かった人文学分野の研究を、自然科学を含む他分野にも共通し得る方法としてのデータ駆動型に再構築し、持続可能な社会構築のためのデータインフラストラクチャを人文学分野に築き、その利活用を通して多分野と協働し得る課題解決型の人文学研究を創成するものである。

古典籍ビッグデータを軸とするデータ駆動型の人文学の研究環境を整備・運営し、人文科学の研究者が多分野と協働して自律的に現代社会にある様々な課題解決に十全に寄与する課題解決型の人文学を創成する。

⑤ 学術的な意義

「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」により構築した大規模画像データベースは、千年以上にわたって蓄積されてきた古典籍データの網羅的な閲覧を可能とした。本計画ではその機械可読テキスト化と検索の高度化を進め、マテリアル解析等の情報等を付加して古典籍ビッグデータを高度化することで非日本語圏や社会科学・自然科学分野に及ぶ広範な活用を可能とする。



図1 プロジェクトの概要

古典籍ビッグデータを用いたデータ駆動型の人文学研究の創成は、世界に先駆けて人文学研究のパラダイムシフトを行う発信源として大きな意義を有している。

⑥ 国内外の研究動向と当該構想の位置付け

今世紀初頭頃からイギリス、フランスを中心としたEU諸国及び米国、韓国、中国などで、それぞれの国や地域に蓄積された古典籍のデジタル化とその公開が国家的事業として急速に進められてきた。人文学研究においても、デジタルヒューマティーズ（DH）に代表される、デジタル化された資料を用いた分析が先駆的研究として行われるようになり、そのための分析ツールなども広く公開されてきてはいるが、こうしたデータの利用と研究は、旧来型の人文学研究の範囲に留まっている。

本計画は、日本のデジタルリソースの活用を、人文学に留まらない多分野に開き、蓄積されてきた古典籍の活用の推進のためのデータインフラストラクチャの構築、ビッグデータのコンテンツ解析・マテリアル分析への提供と得られたデータの環流と蓄積といった、人文学分野と社会科学・自然科学分野との循環的研究環境の構築を通して、データ駆動型の人文学を創成するという、世界の現状に鑑みても先駆的な計画である。

⑦ 社会的価値

古典籍には、我が国に生きた人々の思想と感情の歴史のみならず、対外交流史、農林水産業・工業を含む産業史、災害史、気候変動を含む自然史といったさまざまな分野に関わる事柄が記録されている。こうした経験的〈知〉を含む〈知〉の伝承が千年を越える期間にわたって同一の言語環境における連続したデータとして取得することが可能であることは世界的にも稀なことである。こうした古典籍に由来するデータは、従来は主としてそれに関わる専門研究者にのみ活用されてきたが、本研究において画像データを機械可読化し、それを公開することによって、多くの分野や様々な位相の人々の利活用が可能となる。

⑧ 実施計画等について

<実施計画・スケジュール>

国文学研究資料館を主体として、以下の研究計画を遂行する。

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。（第1～10年次）

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

共同研究第1期（試行的研究）（第1～4年次）

共同研究第2期（開発的研究）（第5～4年次）

共同研究第3期（実装的研究）（第6～10年次）

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。（第1～10年次）

<実施機関と実施体制>

本研究は、国文学研究資料館を実施主体とし、学術協定書及び覚書を交わしている諸機関等とともに遂行するものである。このほか、協定にはよらない形でも地域や民間企業等の多様なステークホルダーと共同で研究、事業等を展開する。

<所要経費>

総経費 54 億円（設備費 1.8 億円、人件費 16.2 億円、運営費等 35.6 億円）

⑨ 連絡先

渡部 泰明（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）