No. 73　（グランドビジョン⑩）

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

文献には、現代社会が解決しなければならない困難な課題に対処するための〈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厖大なデータが、文献情報資源として蓄積されつつある。適切な課題を設定し、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によって、過去の人類の英知を〈総合知〉として現代社会に生かすことが可能になる。

②　ビジョンの内容

文献に含まれる〈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、（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うという面があった。また、良く知られた著名な文献だけを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ような弊害が多々あることも否めない。

文献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厖大に蓄積（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、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。

文献資料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ことができる。また、媒体としての文献は、物質としての特性を持つ。日本は世界的にも希に多くの古典籍が残存しているが、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始まりつつあり、文献資料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。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。適切な課題を設定し、人文学、自然科学の各分野と連携して異分野融合によるデータ駆動型研究を創成することで、〈総合知〉に至る道を切り開くことができると確信している。

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。

歴史的典籍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。

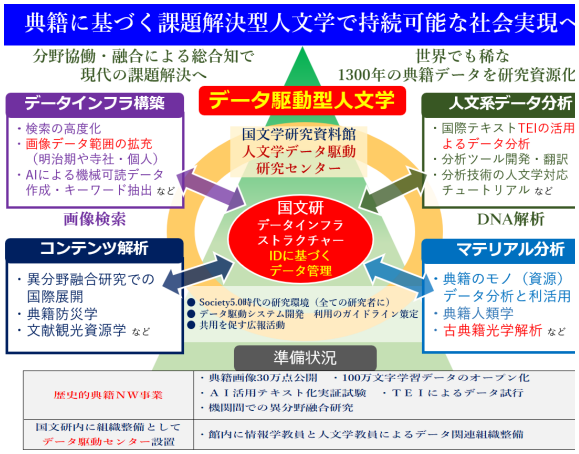

図１　プロジェクトの概要

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」によって構築される大規模画像データベースにより、千年以上にわたって蓄積されてきた歴史的データの網羅的な閲覧が可能となった。本計画ではその機械可読化と検索の高度化を推進し、マテリアル解析等の情報等を付加してデータを高度化することにより、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。

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、世界的に先駆けて人文学研究のパラダイムシフトを行うもの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書物や文書類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、蓄積されてきた歴史的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

日本語の歴史的典籍には、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。こうした歴史的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

以下の４つの研究計画を遂行する。

(1)データインフラストラクチャの構築

データ駆動型システムの開発、運用（維持管理及び改修）、データアーカイブ機能の強化、権利関係等のガイドラインの検討・策定と分野横断的データカタログ整備、データ間の時系列等接続関係の整備

(2)人文学系データ分析技術の開発

メタデータ付与に関する合意形成及び汎用的仕組みの検討と開発、国際テキスト（ＴＥＩ）へのフォーマット作成及び作成ツールの開発、作成ツールの向上化、画像検索・解析技術の向上と可視的把握技術の確立、ＡＩ技術に基づく機械可読データの自動化の開発

(3)コンテンツ解析からの展開

典籍防災学の拡大、典籍人類学の構築

(4)マテリアル分析・解析

マテリアルとしての書物の分析技術の確立、マテリアルとしての書物の復元技術の確立

あわせて、30 万点を超える画像データの拡充を推進し、海外発信・連携機能の強化のため国内外でのシンポジウム等を開催する。また、画像・テキスト等人文系資源に基づくデータ駆動型人文学研究への展開を推進する。

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、情報システム研究機構の協力をあおぎ、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」において学術協定書及び覚書を交わしている諸機関とともに遂行する。

所要経費

総経費 54.2 億円を計上している。内訳は以下のとおり。

設備費（データベースシステム構築・維持管理）３億円

人件費（特任研究員、研究員（ポスドク）、事務職員等）36.4 億円

運営費（一般管理費等）0.8 億円

その他（画像データ作成費、共同研究経費等）14 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）