

要 旨

1 作成の背景

情報技術の急速な進展は、人文学・社会科学の研究・教育の方法を根本から変えつつある。紙資料をはじめとする様々な物理的オブジェクトを基盤として発展してきた従来の手法は依然として重要ではあるものの、研究の資源・成果を持続的な形で保存し、再利用可能な知識資源とするためには、オープンサイエンスを踏まえた研究データ基盤が不可欠である。とりわけ人文学・社会科学研究のための質的データの構築と公開は、現代社会における最重要課題の一つである。質的データには、社会調査におけるインタビュー記録や逐語録、デジタル・ヒューマニティーズ (DH) における構造化テキストデータ、翻刻や注釈を付した史資料データ、画像・音声・動画・3Dモデルを含む多様な文化資料のデータなど多彩なものがある。これらを機械可読性の高い質的データとして流通させることは、生成AIが急速に普及する現時点において、より「正確な」社会文化の理解に不可欠である。

2 現状及び問題点

欧州では 1990 年代末から質的データアーカイブが整備され、オープンサイエンスの実現に向けた潮流の中で FAIR 原則 (Findable, Accessible, Interoperable, Reusable) に基づくデータの保存と再利用の枠組みが着実に確立されてきた。人文学・社会科学分野でも欧州連合の下で各分野の特性に合わせたインフラ事業が欧州全域で展開されている。これにより、各国で研究データ管理の取組みが制度面・運用面の双方から推進され、研究者がデータにアクセスし共有できる環境が大幅に整ってきた。

さらに、これらの研究インフラは関連する国際的なデファクト標準に基づいて構築され、その結果、人文学・社会科学の研究者がデータやコードを公開し再利用可能な形で共有することが、欧州における標準的な研究実践として定着している。

一方、日本では、いくつかの先進的事例は存在するものの、対象分野や地域には偏りがあり数も少なく、包括的基盤には至っていない。構造化テキストデータや注釈付き資料の共有基盤の構築に至ってはようやく着手した段階であり、研究成果の透明性・再利用性・国際接続性も十分に確保されているとは言い難い。また、関係者間での国際標準規格に関する情報共有及び周知も十分ではなく、研究者・専門職の間で理解と実装にばらつきがある。加えて、それらを推進すべき質的データの構築・管理を担う専門人材が不足しており、その育成体制も十分とはいえない。

日本学術振興会が推進する人文学・社会科学に関するデータインフラストラクチャーの事業では、選定された研究機関において構築された研究データのメタデータを整備・共有して発見性を高める取組みが進みつつある。しかし、多くの研究組織・研究者による研究データの構築やそれを踏まえた基盤整備については、いまだにその個別的取組みに依存している。結果として、網羅的な基盤はなく、制度的に支えられた枠組みも存在せず、データの維持管理や長期利用保証にも課題が生まれている。

さらに、日本列島特有の言語文化的要件を国際標準に反映させる仕組みも未整備である。これらの課題を受け、2024 年度より文部科学省が開始した「人文学・社会科学の DX 化に向けた研究開発推進事業」では、DH コンソーシアムの運営が始まり、東アジア／日本のテキスト構造化や地理情報に関するガイドラインの策定が進められている。こうした動きを着実に発展させることが求められる。

社会の多様な声や文化資源を保存・共有し、それを教育や政策形成へと還元する営みは、民主主義の基盤を強化するとともに、SDGs の達成にも寄与する。また、生成 AI 時代にふさわしい知識基盤としても欠かせない要素である。既に始動している取り組みを着実に発展させ、国際標準との連携を図りながら、公開できるデータは可能な限り公開し、個人情報や権利関係等により公開に制約のあるデータは適切な条件の下で利用可能とするという、公開（オープン）と制限付き利用（クローズ）の双方を適切に組み合わせた持続可能なエコシステムを構築することこそ、デジタル時代における日本の人文学・社会科学の確かな発展を支える道筋となる。こうした基盤の整備は、最終的に人類の知の総体（body of knowledge）を一層豊かにする礎ともなる。

3 見解の内容

(1) 制度的に支えられた基盤整備の推進

中核機関を事務局とし、関連する多くの研究教育機関を参画機関とする人文学・社会科学の質的研究のための研究データ基盤に関するコンソーシアム型組織を立ち上げ、ELSI（倫理的・法的・社会的な課題）への対応を踏まえた基盤の構築を推進する。

(2) 持続可能性を担保する中核機関の設置

新たに機関を設立するのではなく、人文学・社会科学研究基盤の共同利用を進める既存組織に「中核機能」を持たせる。常勤職員を十分に配置し、任期付きポストを可能な限り排除して雇用を安定化させることで持続可能性を担保する。この中核機関は、国内の関連事業の成果を基礎として各機関の情報資源を連携させるとともに、取扱いに慎重を要するデータについても適切な処理を経て検索・活用できるシステムを構築し、これを通じて国際的研究インフラとの接続を確保する。

(3) 国際標準への積極的参画

人文学・社会科学に関する様々な国際標準及びそれに類する規格に対し、中核機関を通じて研究者コミュニティが情報を適切に共有しつつ継続的に関与できる仕組みを整え、国際的な関連コミュニティへ組織的に参画することで日本や東アジアの言語文化の特性を国際的枠組みに反映させる。

(4) 教育と人材育成の強化

既存の質的データ作成や利活用を進めるプログラムを有する大学等組織を拠点校とし、修士課程に「質的研究データ基盤論」などの共通科目を設置する。初期段階では修士人材を育成し、研究者・専門職を対象とする短期集中研修も並行して実施する。その後、5 年を目処に全国展開し、博士課程・専門職大学院に拡張する。これを支える教育用データセット及びツールを開発・整備する。