見　　解

人文学・社会科学におけるオープンサイエンスを踏まえた質的研究のためのデータ基盤の形成

令和８年（２０２６年）９月10日

日　本　学　術　会　議
地域研究委員会・言語・文学委員会・哲学委員会
・心理学・教育学委員会・社会学委員会・史学委員会
・法学委員会・経営学委員会・情報学委員会合同
デジタル時代における新しい人文・社会科学に関する分科会

この見解は、日本学術会議地域研究委員会・言語・文学委員会・哲学委員会・心理学・教育学委員会・社会学委員会・史学委員会・法学委員会・経営学委員会・情報学委員会合同デジタル時代における新しい人文・社会科学に関する分科会の審議結果を公表するものである。

日本学術会議地域研究委員会・言語・文学委員会・哲学委員会・心理学・教育学委員会・
社会学委員会・史学委員会・法学委員会・経営学委員会・情報学委員会合同
デジタル時代における新しい人文・社会科学に関する分科会

委員長	永崎　研宣	（連携会員）	慶應義塾大学文学部教授／一般財団法人人文情報学研究所・主席研究員
副委員長	矢野　桂司	（第一部会員）	立命館大学文学部教授
幹　事	平田　貞代	（連携会員）	芝浦工業大学大学院理工学研究科准教授／東北大学大学院工学研究科特任准教授
幹　事	後藤　真	（連携会員）	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准教授・本部機構長戦略室特命准教授
	岩井　紀子	（第一部会員）	日本版総合的社会調査共同研究拠点大阪商業大学JGSS 研究センター長
	大橋　幸泰	（第一部会員）	早稲田大学教育･総合科学学術院教授
	川嶋　四郎	（第一部会員）	同志社大学法学部法律学科教授
	小長谷　有紀	（第一部会員）	武雄アジア大学学長
	坂田　省吾	（第一部会員）	新潟医療福祉大学心理・福祉学部心理健康学科教授
	鈴木　基史	（第一部会員）	京都大学名誉教授
	谷口　尚子	（第一部会員）	慶應義塾大学法学部政治学科教授
	野口　晃弘	（第一部会員）	南山大学経営学部経営学科教授
	原田　範行	（第一部会員）	慶應義塾大学文学部教授
	吉岡　洋	（第一部会員）	京都芸術大学文明哲学研究所教授
	吉水　千鶴子	（第一部会員）	公益財団法人東洋文庫研究部研究員／筑波大学人文社会系名誉教授
	大場　みち子	（第三部会員）	京都橘大学デジタルメディア学部デジタルメディア学科・教授
	小川　健一	（連携会員）	防衛大学校防衛学教育学群統率・戦史教育室教授
	隠岐　さや香	（連携会員）	東京大学大学院教育学研究科教授
	奥田　太郎	（連携会員）	南山大学社会倫理研究所教授
	木部　暢子	（連携会員）	大学共同利用機関法人人間文化研究機構長
	木村　敏明	（連携会員）	東北大学大学院文学研究科研究科長・教授
	金水　敏	（連携会員）	放送大学大阪学習センター所長

瀧川　裕貴	（連携会員）	東京大学大学院人文社会系研究科准教授
橋本　隆子	（連携会員）	千葉商科大学副学長・商経学部教授
日比　嘉高	（連携会員）	名古屋大学大学院人文学研究科教授
松田　陽	（連携会員）	東京大学大学院人文社会系研究科准教授
師　茂樹	（連携会員）	早稲田大学文学学術院教授

本見解の作成に当たり、以下の方に御協力いただいた。

戸谷　圭子	（第一部会員）	明治大学専門職大学院グローバルビジネス研究科教授

本見解の作成に当たり、以下の職員が事務を担当した。

事務局	郷家 康徳	参事官（審議第一担当）
	加瀬 博一	参事官（審議第一担当）付参事官補佐
	高畑 麻衣子	参事官（審議第一担当）付審議専門職付

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。

目　　次

１　本見解の目的と質的データの概念 1
(1)　国際的動向と本見解の目的 1
(2)　質的データの概念と対象 2
(3)　これまでの意思の表出との関係 4
(a)　学術情報流通に関するもの 4
(b)　人文学に関するもの 4
(c)　社会科学に関するもの 4
(d)　人文学・社会科学に関するもの 4
(e)　生成 AI に関するもの 5
２　国際的動向と先行事例 5
(1)　欧州における基盤整備の展開 5
(2)　北米における取組み 6
(3)　国際標準規格の整備と普及 6
３　日本の現状 7
(1)　人文学における基盤的取組み 8
(2)　社会科学における現状 10
(3)　計算社会科学の展開 11
(4)　教育と人材育成の多様化 11
(5)　日本における課題 13
(a)　制度的整備の不足 13
(b)　データ共有に際しての課題 13
(c)　国際標準規格への対応と国際参画の遅れ 14
(d)　人材育成と教育体制の不十分さ 14
(e)　持続可能性の欠如 15
(f)　萌芽的取組みの継続的発展の必要性 15
４　課題に対応した具体策 16
(1)　制度的に支えられた基盤整備の推進 16
(2)　持続可能性を担保する中核機関の設置 16
(3)　国際標準への積極的参画とルール形成 16
(4)　質的データの責任ある共有に向けた体制の整備 17
(5)　教育と人材育成の強化 17
(6)　情報共有の強化と国内外ネットワークの構築 18
(7)　対応策の実現によってもたらされる効果 19
５　質的研究データ基盤における ELSI への対応 19
(1)　基本的視点 19
(2)　倫理的・社会的課題と基盤設計 20

(3)　法的・国際的枠組み 20
(4)　学際研究及び AI との関係 20
(5)　実装とガバナンス 21
６　期待される効果 21
(1)　研究の透明性と再現性の向上 21
(2)　学際的・国際的共同研究の促進 22
(3)　教育と人材育成への貢献 22
(4)　社会への知識共有と政策形成支援 22
(5)　AI 時代における研究基盤としての意義 22
(6)　SDGs 及び公共的知識基盤への貢献 23
７　今後の展望 23
(1)　ロードマップの構築 23
(2)　国際連携における日本の責任と可能性 24
(3)　学術政策との整合と推進 24
(4)　社会に開かれ、社会とともに歩む学術基盤へ 25
８　見解のまとめ 25
＜用語説明＞ 28
＜参考文献＞ 34
＜参考資料１＞審議経過 39
＜参考資料２＞シンポジウム開催 40

1　本見解の目的と質的データの概念

(1)　国際的動向と本見解の目的

情報技術の急速な発展は、人文学・社会科学の研究・教育にこれまでにない変化をもたらしている。欧州や北米においては、研究データを保存・共有し、国際的に接続可能な形で流通させるための基盤が二十年以上にわたり整備されてきた。これらの基盤は単なる技術的支援ではなく、学術政策の中核として位置付けられ、研究の透明性・再現性・持続可能性（※）[1]を制度的に担保するものとなっている。その前提となるのは、オープンのみならずクローズドなデータの適切な扱いも視野に入れた FAIR 原則（Findable, Accessible, Interoperable, Reusable）に基づくオープンサイエンスの枠組みであり、質的・量的を問わず研究データを長期保存し、教育・研究・政策立案に広く活用するための仕組みが整えられている。日本学術会議の提言「第７期科学技術・イノベーション基本計画に向けての提言」[1]においても、こうした国際動向と歩調を合わせることの重要性が説かれている。

しかし日本では、このような人文学・社会科学におけるデータ基盤構築の国際的動向が十分に共有されていない。この状況は、量的研究手法のための研究データ（以下、量的データ）よりも、むしろ質的研究手法に資する研究データ（以下、質的データ）において顕著である。量的データについては、定型的な構造を持つため国際的なメタデータ標準や匿名化の手法が早くから確立しており、日本においても社会調査データアーカイブを中心に整備が進んできた（１章(2)及び３章(2)参照）。これに対し質的データについては、欧州では、人文学における国際標準規格に基づく構造化と、社会科学における匿名化手法や利用条件の標準化とを通じて、相互運用可能な形で保存・公開する取組みが既に制度化されている。一方、日本ではいずれの分野においても研究者個人や一部機関の努力に多くを依存し、基盤的整備に十分結び付いていない。

その結果、日本の研究は二つの側面で不利な立場に置かれている。第一に、国際的な基盤についての情報が共有されないまま独自形式でのデータ整備が進められ、再利用性や国際水準に基づく持続可能性の確保が困難になっている。第二に、国際的な議論に十分関与し得ないことで、日本語・日本文化資料に固有の課題、例えば多漢字・多文字処理、縦書き表記、連綿体、多様な注釈形式などが国際標準に反映されにくい状態に置かれている。これは日本の学術資源をグローバルな知識循環に乗せる上で構造的障害となり、さらに、生成 AI が世界中のデータを基盤に学習する中で、日本の人文学・社会科学の成果が学習資源として取り込まれにくいという問題にもつながっている。

本見解は、国際的な基盤整備の現状を整理し、日本が行うべき基盤整備を明らかにした上で、日本の研究資源を国際的枠組みへと接続する道筋を示すことを目的とする。人文学・社会科学の教育・研究に携わる研究者のみならず、研究支援機関や政策担当者に対しても、情報共有を含む複数の課題を統合的に捉え、基盤整備を推進するための方向性を提示するものである。これは、現在における日本の人文学・社会科学の確かな発展

[1] 本文や脚注における初出で「※」を付した語は、巻末の用語説明に項目がある。

を支える道筋であり、ひいては人類の知識の総体（body of knowledge）をより豊かにする礎となる。

(2) 質的データの概念と対象

本見解における質的データとは、主として非数値的な形式で記録され、その意味の理解が、作成・収集・記録された文脈や、並びに作成者、記録者、研究者及び対象者等の関係に強く依存するデータを指す[2]。これらのデータは、数量化のみでは、個々の事例が持つ意味や文脈を十分に保持できない場合が多い。したがって、その分析においては、文脈情報を踏まえた解釈が中心的な役割を果たす。

このような性格を持つデータは、人文学・社会科学に固有のものではない。自然科学を含むあらゆる分野において、量的分析のみでは十分に扱い得ない情報が存在する。これらは、対象の個別性や観察・記録の条件と強く結び付くため、内容や意味の水準における一律の標準化・構造化が難しい。また、その再利用には、データそのものに加えて、その来歴、生成・収集方法、観察条件、記録・処理過程、権利及び倫理的条件等に関する情報が必要となる。この意味において、文脈依存性の高いデータの保存・共有・利活用は、複数の学術分野にまたがる課題である。

ただし、「質的データ」という概念が各分野で指し示す範囲や、研究方法論上の位置付は、分野ごとに大きく異なる。そのため本見解では、人文学・社会科学における質的データに焦点を当てることとし、他分野における類似のデータとの概念上及び方法論上の共通点・相違点については、今後更に検討すべき課題とする。

以上を踏まえた上で、本見解で扱う質的データは、人文学・社会科学において、記述・言語・映像、並びにそれらの成立状況や相互関係に関する情報を手がかりとして収集・分析される研究データ全般を指すものとする。すなわち、数値化してそのまま量的に処理することが容易ではないデータである。この種のデータは技術の発展につれて様々な形式で活用可能となり、研究自体の幅を広げるものでもある。それゆえここでは、これを厳密に定義するのではなく、その広がりの可能性を含むデータ形式の全体として捉え、現在の姿を対象として扱う。

具体的には、社会科学におけるインタビュー記録や逐語録、現地フィールドノート、人文学における歴史資料のデータ（翻刻テキスト・画像）、文学の校訂テキスト・学術編集版（※）、歴史的・文化的資料の目録、文化人類学・民俗学における各種記録類などが典型例である。デジタル・ヒューマニティーズ（※）（以下、DH）の領域では、TEI（Text Encoding Initiative）ガイドライン（※）などに基づき論理構造や注釈を付与した構造化テキスト、画像・音声・動画と結び付いたデータ、三次元モデル等もまた質的データの重要な一部を構成している。

---

[2] 例えば、英国の社会科学研究データ基盤であるUK Data Service（※）では、インタビューの逐語録、日記、フィールドノート、自由記述、音声・映像、画像等の非数値的情報を質的データとしている。

質的データに共通するのは、それが研究者の解釈と記述を不可避的に含み、対象の文脈を重視する点にある。したがってその保存・共有には、単なるデジタル化にとどまらず、解釈過程を含めて記録し利用者に提示する仕組みが必要となる。

ここで対比される量的データとは、対象の属性や行動を数値として測定・集計し、統計的処理を前提として構造化された研究データを指す。社会科学においては、質問紙調査・世論調査・パネル調査等の個票データ、国勢調査をはじめとする公的統計、経済統計や企業・家計に関する数値データ、実験や心理測定に基づく尺度データなどが典型である。人文学においても、コーパスにおける語彙・文法項目の頻度データ、計量文献学における文体統計、歴史人口学や経済史における人口・物価・生産量等の数量データ、考古学における遺物の計測データ、書誌・目録情報の集計など、量的データは少なからず存在する。ただし、人文学における量的データの多くは、翻刻テキスト・史資料・目録といった質的データから導出されるものである。

量的データは、調査の個票、時系列や集計表の形をとる統計、計測機器や実験から得られる測定値、コーパスや文献から抽出された頻度や分布、位置情報や関係を数値で表したネットワークのデータなど、形態は多様である。しかしいずれも、あらかじめ定められた測定・分類・符号化の手続きを経て数値やカテゴリとして記録され、その意味が変数の定義、単位、抽出・測定の方法といった手続き的な情報によって規定される点で共通する。このため、量的データは、社会調査データにおけるDDIのようなメタデータ標準や分野ごとに整備された記述規則の下で機械可読な形に整えやすく、集計・符号化・変数の削除といった匿名化の手法や、利用条件の設定についても定型的な方法が確立されてきた。これに対し質的データは、前述のように意味の理解が文脈に依存するため、構造化と匿名化の双方において、量的データとは異なる方法と制度設計を必要とする。欧州においてはこの要請に応える基盤が人文学・社会科学の双方で整備されてきたのに対し、日本ではそれがなお欠けている。本見解が質的データに焦点を当てる理由はここにあり、これは量的データの基盤整備が完了したことを意味するものではない。

量的データについても、社会調査に関する提言[2]が、公的統計を含む個票データの二次利用が過度に制約されていること、秘匿処置を施した上での個票データ利用環境の構築、社会調査を担う人材の育成を課題として挙げ、社会的モニタリングとアーカイブに関する提言[3]が、データ収集・維持の長期的な保証を欠き先細りにある状況を指摘した上で、アーカイブ化と公開の仕組みを同時に整える必要を説いているように、なお多くの課題が残されている。さらに人文学においては、量的データを集約・共有する基盤そのものが乏しい。本見解は、量的データに関するこれらの課題が既往の提言において示されていることを踏まえ、日本学術会議においてこれまで十分に検討されてこなかった質的データに対象を絞るものである（１章(3)参照）。もとより、４章で述べる基盤は、既存の量的データのアーカイブと接続し、両者を横断的に活用できることを前提として構想される。質的データは量的データと対立するものではなく、むしろ相補的に機能する。

(3)　これまでの意思の表出との関係

日本学術会議は、研究情報流通基盤の整備について複数回にわたり提言を公表してきた。社会科学では研究データに関する提言が着実に重ねられ一定の方向性が示されてきた一方、人文学ではデジタル基盤の整備を求める提言はあるものの、その後の技術や国際標準規格の急速な進展を踏まえた具体的検討は十分ではない。とりわけ、人文学・社会科学に特有の質的研究データ基盤の在り方に即した議論はなお不十分である。以下、関連する提言をテーマ別に確認する。

(a)　学術情報流通に関するもの

学術全般を対象とする近年の包括的提言として、「オープンサイエンスの深化と推進に向けて」[4]（2020 年５月）が研究情報流通基盤の制度整備を、「研究ＤＸの推進」に関する回答[5]（2022 年 12 月）がオープンサイエンスとデータ利活用の推進を、「第７期科学技術・イノベーション基本計画に向けての提言」[1]（2024 年 11 月）が知識基盤としてのデータ基盤・デジタルアーカイブの拡充と高度人材育成を、それぞれ求めている。本見解は、これらが示す方向性を、人文学・社会科学の質的研究データ基盤の形成を通じて具体化するものである。なお、「学術情報流通の大変革時代に向けた学術情報環境の再構築と国際競争力強化」[6]（2020 年９月）は理工系の学術出版を主対象とし本見解とは文脈を異にするが、J-STAGE の活用拡大などに見られる方向性の一部を共有する。

(b)　人文学に関するもの

「学術の総合的発展をめざして―人文学・社会科学からの提言―」[7]（2017 年６月）は、国家的諸機関・図書館・大学・学協会が結集した大規模かつ長期的なデータベース整備の必要性を説いており、本見解もこの主張を改めて確認する。

(c)　社会科学に関するもの

社会調査・Web 調査に関する提言[2]（2017 年９月）・[8]（2020 年７月）は、調査環境の変化や Web 調査の興隆に対応した方法論を示すとともに、取扱いに慎重を要するデータの倫理的配慮や手続きの透明性という、本見解の基盤設計にも通じる論点を含む。また、「社会的モニタリングとアーカイブ」[3]（2020 年９月）が説く総合的アーカイブの必要性は本見解の目指すところと重なり、可視化に関する報告[9]（2020 年 9 月）は、質的研究データ基盤の成果を新たな研究へと展開する上で示唆を与える。なお、これらの提言が示した量的データをめぐる課題認識を本見解は前提とするものであり、量的データの基盤整備が完了したとの認識に立つものではない。

(d)　人文学・社会科学に関するもの

アジア研究を対象とする「新たな情報化時代の人文学的アジア研究に向けて」[10]（2017 年９月）は、海外で構築されたデータベースへのアクセスが、利用する側からの発信を条件とする相互主義的な形で想定されることが多いため、データベース整備の遅れがその国・地域の研究成果の発信力を弱め、研究情報へのアクセスの不平等をもたらすと指摘している。この相互主義的な条件はアジア研究に固有のものではなく、研究資源が複数の国・地域に分散して所在し、国境を越えた利用が不可欠な分野に広

く共通する。本見解が目指す質的研究データ基盤の形成は、日本の研究資源を国際的に相互運用可能な形で発信できるようにするものであり、そこで指摘された課題の部分的な解消にも資する。

(e) 生成 AI に関するもの

「生成 AI を受容・活用する社会の実現に向けて」[11]（2025 年２月）は、生成 AI の基盤として質の高いデータ管理と信頼性の確保が必要であると説いており、質的研究データ基盤の整備・運用はこの方向に沿うものである。

２ 国際的動向と先行事例

(1) 欧州における基盤整備の展開

欧州は、質的研究データ基盤の制度化において世界的に先行している。1990 年代末以降、欧州連合では ERIC（欧州研究インフラ・コンソーシアム）[12]の枠組みの下、人文学研究のデジタル基盤を担う DARIAH（※）、言語資源を集約する CLARIN（※）、社会科学データを統合的に提供する CESSDA[13]が中核的な研究インフラとして体系的に整備されてきた（各基盤の詳細は用語説明を参照）。これらはいずれも ESFRI（European Strategy Forum on Research Infrastructure）ロードマップに登録された「欧州研究インフラ」として制度的に保証され、恒常的な資金と人材配置の下で運営されている。すなわち欧州の基盤整備は、個別研究者の努力に依存するのではなく、国家間の合意と政策に基づく制度として確立されている点に特徴がある。

これと並行して、各国でも国家レベルの整備が進む。DARIAH と CLARIN を国内・地域レベルで統合運用する枠組み CLARIAH がオランダ・ベルギー等で整備され、ドイツでは協働編集基盤 TextGrid（※）が DARIAH-DE を経て、政府主導の全国的研究データ基盤 NFDI[14]の下の Text+や NFDI4Culture 等へと発展している。フランスでは国家研究基盤 Huma-Num（※）が研究データの保存・共有・アクセス支援を包括的に担う。社会科学分野では、CESSDA が 2017 年に ERIC 傘下となり、各国のデータアーカイブ（英国の UK Data Service、ドイツの GESIS、オランダの DANS 等）を Service Provider として束ね、量的データのみならず質的・混合型データを共通メタデータモデルの下で横断的に検索・利用できる仕組みを構築している。とりわけ英国の Qualidata やフィンランドの FSD など、1990 年代から質的データを保存・公開してきた取組みが、欧州規模の共有を支えている。

研究データの公開・流通を支える仕組みも整っている。個々の研究者の研究データを公開する場として Zenodo[15]が広く利用され、Journal of Open Humanities Data[16]や Research Data Journal for the Humanities and Social Sciences[17]の刊行により、研究データの作成・公開を学術的成果として評価する動きも進む。複数の研究インフラが連携する SSH Open Marketplace[18]は、研究資源やツールのメタデータを集約し横断検索を可能にしている。教育面でも、大学院での DH 専攻の設置、Digital Humanities Oxford Summer School（DHOXSS）[19]等のリカレント教育、DARIAH のオンライン教材 dariahTeach[20]の公開など、知識を共有する環境が整備されている。

加えて、こうした基盤整備は技術的・制度的条件の整備にとどまらず、倫理的・法的・社会的課題（ELSI）への対応を前提として進められてきた。質的研究データには、個人情報をはじめとする取扱いに慎重を要する情報を含む社会調査資料に加え、来歴や帰属をめぐって論点を有する文化財・歴史資料も含まれ得ることから、その整備・公開・利活用はELSIと不可分だからである。

(2) 北米における取組み

北米でも、研究データ基盤の整備は早くから進展してきた。米国では、社会科学分野の研究データ基盤として ICPSR（Inter-university Consortium for Political and Social Research）[21]が1962年に設立され、世界各国の社会調査データを収集・保存し、学術研究における二次分析のために提供している。ICPSR は現在、世界最大級の社会科学データアーカイブとして、量的データだけでなく質的調査データや混合型データの保存・提供も進めている。

また、人文学分野では、大学図書館の連携によって構築されたHathiTrustデジタル図書館（HTDL）が、北米を中心とした大規模デジタル資料の保存と共有を担う基盤として運用されている。HathiTrustは数百万冊規模のデジタル化資料を収集し、横断検索を可能にするほか、その資料群を研究用途で分析可能な形で提供する HathiTrust 研究センター（HTRC）(※)を通じて、大規模デジタル資料を用いた計算機的研究を支える環境を提供している。

こうした研究基盤の整備は政策的支援によっても推進されてきた。米国では、全米人文学基金（NEH）が2008年にOffice of Digital Humanitiesを設置し、DHに関する研究プロジェクトやツール開発への資金提供を本格的に開始した。これにより、デジタル資料の活用や研究データの公開を前提とした研究活動が広く促進される方向性が整備された。

さらに、欧州同様、教育・人材育成も積極的に進められている。大学院教育では、研究データの管理や公開、国際標準規格の利用などを研究活動の初期段階から学ぶカリキュラムが整備されている。代表的な例として、カナダのヴィクトリア大学が長らく主催し、現在はモントリオール大学に拠点を置く Digital Humanities Summer Institute（DHSI）[22]が挙げられる。DHSIは世界各国の研究者や大学院生が参加する国際的な教育プログラムとして広く知られ、デジタル資料の分析手法や研究データの管理・公開に関する実践的な教育を提供している。

(3) 国際標準規格の整備と普及

研究データの国際的共有を可能にしているのは、国際標準規格の存在である。現在広く利用されている国際標準規格としては、文字情報については Unicode、テキスト構造化についてはTEIガイドライン、画像の相互運用についてはIIIF（International Image Interoperability Framework）[23]、社会調査データについてはDDI（Data Documentation Initiative）[24]がある。これらの標準は、研究成果を機械可読な形で保存し、国際的に流通させるための前提条件となっている。

とりわけ TEI ガイドラインは、翻刻や注釈を伴う質的データを国際的に共有可能な形で表現するための基盤として、主に欧米の人文学分野で広く活用されている。例えば、英文学研究における大規模なものとしては、1470 年から 1700 年の間の英語作品を扱う Early English Books Online(EEBO)や 18 世紀の刊行物を扱う Eighteenth Century Collections Online(ECCO)といった図書資料の大規模画像データベースに対して主に英語圏の国際的な研究図書館 200 館の連携によってテキストデータ化とその公開に取り組んだ TCP(Text Creation Partnership)[25]がある。TCP は、EEBO 及び ECCO に収録された資料の中からテキスト化可能なものを選定し、TEI ガイドラインに準拠した構造化テキストとして公開してきた。具体的には、EEBO では約６万点、ECCO では 2,231 点が TEI 準拠データとして整備され、これらは自由に利用可能なデータとして提供されている。この公開により、各地で検索・閲覧サイトが構築されるほか、多様な研究で活用されるなど、DH における基盤的資源として広く位置付けられている。ECCO-TCP のオープンデータ化に際しては、商業学術データベース出版社との協働により実施された。英語学では、１億語の単語への品詞情報を含むイギリス英語のコーパスである British National Corpus[26]も TEI 準拠データとして無償公開されている。また、多言語にまたがる取組みとしても、例えば、Distant Reading for European Literary History(COST Action CA16204)[27]では、計算文学研究分野の研究者達が集い、14 言語で 2,000 冊以上の小説の全文テキストデータを TEI ガイドライン準拠で作成し公開している。グレコ・ローマ時代の古代から 20 世紀に至るまでの主に欧州の演劇を計算機的に研究するために開発された DraCor（“drama corpora”の略）[28]では、22 言語、4,000 点以上の、戯曲に特化する形で TEI 準拠で構造化されたテキストが公開され、様々な活用が可能となっている。このように、TEI ガイドラインは、テキストデータを広く共有して活用されることを目指すプロジェクトに採用されている。

TEI ガイドラインを含め、規格の改訂には研究者自身が参画することが可能であり、国際的コミュニティの活動に加わることが、自国の研究資源を世界に発信することに資する重要なプロセスとなっている。とりわけ、国際標準規格への対応は、相互運用性や流通可能性を確保するための技術的課題にとどまるものではない。質的研究データを標準化し、広く共有・再利用可能な形で整備することは、個人情報をはじめとする取扱いに慎重を要する情報の処理、資料の来歴や帰属をめぐる論点にも容易に接続し得る。また、文化的・言語的多様性を包摂するためには国際標準規格としての対応は必須となる。さらに、国際標準規格に準拠した質的研究データは AI による利用可能性を高めることにつながりやすい。そうした様々なテーマは、倫理的・法的・社会的課題（ELSI）とも深く関わる。このため、国際標準規格への対応やその改訂への参画は、ELSI を踏まえた質的研究データ基盤の設計と運用の一環として位置付け得るものである。

３　日本の現状

これら国際的な状況と連携の課題を踏まえつつ、日本の現状を整理する。なお、以下で見るように、質的データ基盤の課題は、人文学と社会科学とで異なる形で現れる。この相

違は、両分野の研究対象と方法論の違いに由来する。人文学においては、テキスト・画像・史資料といった研究対象そのものが質的データであり、図書館・文書館・博物館等による資料のデジタル化、翻刻、目録作成という形で、質的データの作成は研究活動と一体のものとして早くから進められてきた。他方、人文学における量的データは、そうした質的データから派生的に作成されるものが多く、量的データを独立に集約・共有する基盤への需要は相対的に小さかった。そのため人文学における課題は、既に大量に作成されている質的データが、国際標準に基づく構造化と相互運用性、及びそれを支える制度的基盤を欠いたまま孤立的に蓄積されている点に集中する。

これに対し社会科学においては、質問紙調査等に基づく量的研究が主要な方法論の一つとして確立し、個票データの二次分析への需要が高かったことから、量的調査データのアーカイブ化が先行した。一方、インタビュー記録、逐語録、フィールドノート等の質的データは、個人の特定につながる情報を文脈の中に含み、また調査者と対象者の信頼関係に基づいて収集されることから、量的データと同じ枠組みでは共有できない。そのため、社会科学における課題は、質的データを責任を持って共有するための制度設計そのものが手つかずであるという点にある。（３章(2)をも参照）

このように、人文学においては構造化と相互運用性、社会科学においては責任ある共有の制度設計が、それぞれ質的データ基盤の課題の中心となる。両者は同一の課題の二つの側面であり、国際標準への対応（４章(3)）と責任ある共有の体制（４章(4)）を併せ持つ基盤によって初めて応えることができる。本見解が両分野を一括して扱う理由はここにある。

(1)　人文学における基盤的取組み

人文学分野においては、研究対象であるテキスト・画像・史資料そのものが質的データであることから、量的データに比して質的データの基盤整備が先行して進められてきた。人間文化研究機構国立国語研究所の「現代日本語書き言葉均衡コーパス」[29]や「国語研日本語ウェブコーパス[30]」は、日本語研究にとって不可欠な基盤であり、国際比較研究にも資する大規模リソースである。これらのコーパスは計量的な言語研究に広く用いられるが、それ自体は形態論情報等を付与した構造化テキストであり、量的分析の基礎となる質的データの典型例でもある。東京大学史料編纂所のデータベース検索サービス（SHIPS）[31]は、一次史料のテキストと刊行された書籍の画像を結び付けて提供することで、歴史資料研究の情報発見と読解作業を飛躍的に高速化させた。1994 年に発足した SAT 大蔵経テキストデータベース研究会[32]は、仏教典籍を IIIF 対応の画像と校正済み構造化テキストデータとして構築した上で両者を連携させて国際的に利用可能な形で仏典データベースとして公開しており、2025 年度より科学研究費・特別推進研究の助成を受けて更なる基盤整備が進められている。

人間文化研究機構国文学研究資料館が運用する「国書データベース」は、『国書総目録』を継承した同資料館の基礎事業である『古典籍総合目録データベース』編纂事業を基盤として発展し、2014 年度から 2023 年度まで実施されてきた「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」を通じて、古典籍研究を支える中核的データ基

盤として整備されてきた。2023 年には、「新日本古典籍総合データベース」と「日本古典籍総合目録データベース」が統合され、書誌情報と画像を一体的に扱う全国的な古典籍データベース「国書データベース」[33]が形成され、現在では約 30 万点の画像が公開されている。IIIF や DOI の採用、クリエイティブ・コモンズによるライセンス付与など、国際的に通用する技術規格と利用条件表示に基づく公開体制も整備されている。現在は、文部科学省の大規模学術フロンティア促進事業「データ駆動による課題解決型人文学の創成プロジェクト事業」を通じて、国内各機関の所蔵資料を連携させる全国的な共同利用の枠組みが構築されつつあり、同データベースの拡張と高度化に寄与している。

立命館大学アート・リサーチセンター（ARC）[34]は、文化資源のデジタル保存・公開を推進し、2019 年には文部科学省から「国際共同利用・共同研究拠点」に認定された。ARC は浮世絵や古地図、映画資料等、多様な日本文化資料に関する多様なデジタルアーカイブを整備し、クラウド基盤を用いた国際的共同研究を展開している。美術史・地理学などの人文学データ基盤が国際的に共有・活用され得ることを示す重要なモデルである。

加えて、日本研究に資するデジタルデータの提供という観点では、ジャパンサーチによる国内デジタル文化資料の横断検索ポータル、そして同ポータルにも収録される国立国会図書館デジタルコレクションにおける大規模なデジタル化資料公開が重要な役割を果たしている。これらは研究の当事者が研究成果のために作成する研究データとは異なり、資料を対象として一括作成されたものであり、特定の研究を念頭に置いたものではないが、デジタルデータとして研究に資するものであり、その意味では学術研究に対する貢献は大きい。国立国会図書館次世代システム開発研究室[35]が技術開発を牽引してきたことにより、国立国会図書館デジタルコレクションでは全文検索機能が整備され、デジタル化された頁画像のみならず、OCR を用いてそこから生成したテキストデータやそれを対象とした検索・表示システムも公開されており、研究データとして活用可能な側面を更に拡張し得る基盤となっている。また、2022 年に開始された個人向けデジタル化資料送信サービスは、絶版等により入手が困難であることが確認された 150 万点以上の資料を、個人認証を条件として広く提供しており、日本研究に対する貢献は極めて大きい。ただし、このサービスは日本国外からの利用が認められていない。

また、IIIF に準拠したデジタルアーカイブが国立国会図書館や国文学研究資料館で公開され、さらに大学図書館にも徐々に普及しつつあることにより、これらの資料を基盤とした質的研究データの作成が進み、機械可読性が高く国際連携にも対応可能なデータが着実に増加している点は、極めて重要な進展である。しかし、このような研究データの持続可能性を担保する上では、情報資源を統一的に識別するための URI の永続性が不可欠である[36]が、システム更新の際に URI を変更してしまう事例が、複数の国立大学図書館においても発生している。

個々の研究者が研究データを公開・共有しようとした場合には、J-STAGE Data において論文の根拠データとして論文とともに公開することが可能であり、さらに、日本デジタル・ヒューマニティーズ学会ではこれを利用したデータ論文掲載の仕組みを同学会誌

[37]において提供しており、既に数本のデータ論文が掲載されている。各大学の機関リポジトリにおいても、研究データを受け付けるところが出てきている。しかしながら、これらは人文学・社会科学におけるデータの特性を十分に活かした形のリポジトリとはなっておらず、公開されるデータも特にフォーマットが定められているわけではなく、再利用はそれほど容易ではないという状況である。

以上のように、課題はありつつも、公開・共有の手段は整いつつある。とはいえ、研究者が個々に作成する翻刻データや注釈付き資料の多くは、なお孤立的に整備され、国際的な相互運用性を備えないまま蓄積されているのが実情である。これを国際的に共有可能な形で構造化するための基盤や標準規格に関する情報が、研究者に十分浸透していないことがその背景にある。

一方、質的研究データに関する個別の取組みを連携させようとする流れも少しずつ形成されつつある。先述した日本学術振興会の「人文学・社会科学データインフラストラクチャー構築推進事業」及びそれに続く同「強化事業」（以下、データインフラ事業）では、関連の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を進めている。2024 年度から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（DiHuCo）（以下、人文学 DX 事業）」[38]は、国内の分野横断的な取組みを推進する第一歩である。同事業では、DH コンソーシアムが運営され、東アジア／日本のテキスト構造化及び地理情報に関するガイドライン制定が進められている。このガイドラインは、TEI ガイドライン等の国際標準規格を東アジア・日本の資料に適用する際の指針として研究者コミュニティが自律的に策定するものであり、国家間の協議や合意を要するものではない。しかし、漢文資料をはじめ東アジア全域で共有されてきたテキストを対象とする以上、日本のみで独自の方式を定めれば、地域内でのデータ共有を損なうことになる。そのため策定に当たっては、TEI 協会の東アジア／日本語分科会などを通じて東アジア各地の研究者と協議しつつ、国際標準規格の枠内で整合性を保つことを前提としている。このような取組みは、研究者コミュニティが主体となって運営される国際標準規格の形成過程に、東アジアの研究者と連携しつつ日本が参画するための重要な土台となる。これらの事業の成果が、短期的に消尽されることなく、研究者コミュニティに広く共有され、かつ、長期にわたり研究基盤として安定的に運用されることが期待される。

(2) 社会科学における現状

一方、社会科学分野においては、量的研究が主要な方法論の一つとして確立し、個票データの二次分析への需要が高いこともあって、量的調査データの整備と公開が質的データと比較して進展している。東京大学社会科学研究所が運営する「SSJ データアーカイブ」は、日本における代表的な社会調査データアーカイブであり、個票データを研究者が二次利用可能な形で提供している。さらに、上記データインフラ事業には、SSJ データアーカイブに加えて大阪商業大学の JGSS 研究センター、一橋大学経済研究所、慶應義塾大学経済学部附属経済研究所パネルデータ設計・解析センター等の尽力により、連携したデータ検索を実現できている。

しかし、質的調査データの整備は依然として強く求められる。インタビュー記録、逐語録、フィールドノートなど、社会調査の厚みを支える基礎資料は、個人の特定につながる情報を文脈の中に含むため、量的データに用いられる定型的な匿名化になじまず、また、調査者と対象者の信頼関係に基づいて収集されることから、量的データと同じ枠組みでアーカイブすることができない。国際的にはICPSRを先例とするデータアーカイブの伝統とDDI等の標準が量的データについて早くから存在したのに対し、質的データについてはこうした固有の性質に即した別個の枠組みが必要とされてきた。欧州においては1990年代末から、質的データに固有の匿名化手法の標準化、利用者資格や利用目的に応じた利用条件の階層化、再利用を見据えた同意文書の設計、データ寄託時の支援体制といった制度的枠組みを通じて、質的データの再利用が制度化されてきた。これらの枠組みは、量的データのアーカイブの単純な延長ではなく、質的データの性質に即して設計されたものである。日本にはこれに相当する枠組みがなく、質的調査データは研究者個人や研究グループ内にとどまり、研究者の退職や異動とともに失われる場合も少なくない。質的調査データを受け入れ、匿名化を支援し、条件付きで提供する体制は、量的データのアーカイブが整備された現在においても、なお萌芽的段階にあるといえる。

(3) 計算社会科学の展開

近年、日本においても計算社会科学（CSS）の取組みが着実に広がりつつある。計算社会科学は、量的データ解析に加えて、量的処理を施す前のテキストや画像・動画データの質的な読解を伴うことが多く、量的解析の妥当性を最終的に検証する上でも質的データの直接的な検討を要するため、質的データを保存・共有する基盤が不可欠である。神戸大学計算社会科学研究センター（CCSS）[39]は、データベース構築・シミュレーション・可視化を三本柱として研究と人材育成を推進し、国際誌Journal of Computational Social Science[40]を発行するなど、国際的発信拠点として重要な役割を果たしている。Summer Institute in Computational Social Science[41]のサテライト・ワークショップSICSS Tokyo[42]では、SNSデータ解析、ネットワーク分析、シミュレーションなど、国際的に共有された方法論を若手研究者へ提供する枠組みが整えられている。さらに、東北大学文学研究科では、人文学と社会科学の横断的連携を視野に入れた大学院教育の一環として計算人文社会学分野[43]が設置され、教育体制が拡充されている。

しかしながら、これらCSS分野の研究成果や教育機会が、研究コミュニティ全体に十分に行き渡っているとは言い難い。質的データの保存・公開をめぐる国際的な議論との接続や、計算社会科学が生成するモデル・ログデータ・シミュレーション結果をどのように研究データ基盤に取り込み、長期的に保存・再利用可能な形で位置付けていくのかは、今後の重要な検討課題である。

(4) 教育と人材育成の多様化

データ基盤の構築と国際的な利活用を促進するための基礎としてはDHの教育が位置付けられる。日本の大学におけるDH教育は、多様な展開を見せている。東京大学大学院人文社会系研究科の次世代人文学開発センター人文情報学部門[44]は、人文学と情報学を架橋し、構造化テキストや自然言語処理を取り入れた大学院教育を推進している。立

命館大学大学院文学研究科の文化情報学専修[45]は、立命館大学アート・リサーチセンターと連携して実践的教育を行ってきた。同志社大学文化情報学部・研究科[46]は、文学・考古学・芸術など幅広い文化資源をデータ科学的に分析する教育を展開しており、学部から大学院まで一貫した人材養成を行っている。大阪大学大学院人文学研究科では人文学林[47]を設置して DH 教育への本格的な取組みを開始した。東北大学の文学研究科における計算人文社会学分野については上述の通りである。

さらに、文部科学省の「デジタルと掛けるダブルメジャー大学院教育構築事業」[48]に基づき、九州大学の「人文情報学大学院（人文情報連係学府）」[49]が設置され、人文学と情報科学を統合的に学ぶ修士課程教育が開始された。千葉大学を中心とした複数大学の大学院で構成する卓越大学院プログラム「アジアユーラシア・グローバルリーダー養成のための臨床人文学教育プログラム」[50]でも、Digital Humanities 2.0 科目として DH を学ぶことが求められている。卓越大学院プログラムは 2025 年度で終了したが、今後、更に連携する大学院を増やしつつ教育体制を充実させることが求められる。これらの事例のように、政策的支援に基づく教育体制の整備は、学際的な研究人材育成を加速する新たな流れとなっている。

上記にとどまらず、DH や歴史情報学等、質的研究データを中心的に扱う分野について、カリキュラムとして整備されるには至っていないものの、個別の授業として提供されている例もある。筑波大学や岡山大学、慶應義塾大学、総合研究大学院大学等では、大学院生がこれに関連する科目を選択することが可能となっている。

しかしながら、このような様々な取組みが進みつつあるにも関わらず、人文学・社会科学分野における多様な研究方法論を前提とした場合には、現時点では各々の方法論を十分に活かす形での教育を実施できているわけではない。あくまでも限定的なものとなっており、これをよりきめ細かに対応させていくことでそれぞれの研究方法論を踏まえたデジタル技術の活用手法を教育できるようにすることが必要である。そうした取組みを通じ、質的データの構築や活用はその意義や有用性をより深めていくことができる。

教育体制の整備に当たっては、学習段階に合わせて質的研究データを効果的に活用できるツールとデータセットを含む教材が必要である。これは人間文化研究機構における DH ツール開発の公募[51]や、東アジア古典籍向けの TEI 対応ビューワ[52]、構造化テキストデータ構築の入門書及びそれを対象とした分析用プログラムのチュートリアル[53]など、一部で作成が進められつつあるものの、ようやく緒に就いたという段階といわざるを得ない。今後これらの動きを拡充しつつ、持続的に運用する構造としていく必要がある。

図書館・博物館・文書館等における専門職もまた、これらの整備には重要な役割の一端を担うはずであるが、現状ではまだ十分な研修体制が整備されているとはいえない。近年、海外の日本研究司書のコミュニティにおいては北米の NCC(The North American Coordinating Council on Japanese Library Resources)[54]や欧州の EAJRS(European Association of Japanese Resource Specialists)[55]が年次総会においてこうした課

題についての研修を開催するようになっている。このような動きについても、海外の日本研究支援という観点から支援・拡充する必要がある。

また、上述の「人文学 DX 事業」でも、人材育成に関する取組みも推進しており、モデルになり得るカリキュラムの策定に取り組んでいる。ただし、この取組みも萌芽的段階であるため、更なる拡充が求められる。

(5) 日本における課題

これらの国際的動向と日本の現状を踏まえると、日本においては複数の先進的取組みが存在するものの、それらは制度的・構造的観点から見た場合、分野横断的基盤として十分に統合されているとは言い難い。国際標準規格や研究基盤に関する情報が研究者及び専門職の間で十分に共有されておらず、結果として実装にばらつきが生じている。

日本の人文学・社会科学・計算社会科学の各分野においては、国文学研究資料館による国書データベース、立命館大学アート・リサーチセンターの国際拠点化、社会科学分野における SSJ データアーカイブなど、先進的な取組みが既に存在している。また、各大学院における教育プログラムなども含め、研究者や学生の間で活用されている実践も少なくない。

しかしながら、これらの取組みの多くは個別機関や研究プロジェクトに依存しており、分野横断的に質的研究データを共有・活用するための制度基盤の整備はなお途上にある。量的データについては、社会科学を中心に一定程度の情報共有が進展している一方、質的データについては、人文学・社会科学のいずれにおいても、データ作成や共有に関わる基盤的情報が十分に蓄積・流通しておらず、依然として発展途上の段階にある。ただし、３章冒頭で述べたとおり、人文学では国際標準に基づく構造化と相互運用性に関する情報が、社会科学では責任ある共有の制度設計に関する情報が、それぞれ欠けているという違いがある。こうした状況を踏まえ、日本における質的研究データ基盤整備の課題として、以下の点を指摘できる。

(a) 制度的整備の不足

現在の取組みは主として個別機関及び研究プロジェクト単位で実施されており、人間文化研究機構国立国語研究所や東京大学史料編纂所などの先進的事例は存在するものの、分野横断的に機能する制度的基盤は未整備である。

欧州では先述の通り、DARIAH や CLARIN、CESSDA などの研究インフラが整備されており、ドイツでは NFDI の枠組みの下で人文学・社会科学分野のコンソーシアムが形成されている。このようなコンソーシアムでは、国家的枠組みの下で研究データ基盤が持続的に運営されている。また、NFDI には年間最大 9,000 万ユーロ規模の資金が投入され、長期的な制度基盤として機能している。

これに対し、日本では国家的イニシアティブとして持続的に支援される仕組みが十分に整っていないことが、質的研究データ基盤整備の大きな課題となっている。

(b) データ共有に際しての課題

オープンサイエンスを前提とした質的データの共有においては、オープン／クローズ戦略を踏まえ、FAIR 原則に基づいた運用が必要となる。というのは、データは必ず

しも公開可能なものばかりではなく、公開不可能なデータも少なからず存在しており、公開可能だとしても一定の制約を必要とする場合もある。例えばインタビューデータの場合にはデータの公開が不可能な場合も少なくなく、このような場合には何らかの条件の下で研究利用を可能とする枠組みが必要となる。また、近世被差別部落が記載される古地図・古絵図の扱いについては学術活動を前提とした公開への道筋は示されているものの、相当な配慮が必要であるとされている[56]。一方、質的研究データの対象として扱われる文化財や歴史資料の中には、その来歴や帰属をめぐって歴史的・倫理的・政治的な論点を伴うものがある。とりわけ、植民地支配や戦争、占領などの非対称な権力関係の下で、略奪・接収・不当な取引等により本来の所在地や帰属共同体のもとから他国の機関等の手に渡った資料については、現在の所蔵者による質的研究データとしての整備や公開それ自体が、その帰属や返還をめぐる国際的な論争と無関係ではあり得ない。日本が収蔵する資料についても、同様の課題が存在し得ることには留意する必要がある。このように、著作権保護期間が満了している資料の共有に関しても分野や資料の性質に応じて様々な課題が存在する。したがって、各分野の専門家による情報の集約と対応方針の検討を幅広く行う必要があり、新たに生じる課題についても同様に取り組んでいく必要がある。

(c) 国際標準規格への対応と国際参画の遅れ

質的データの整備に当たり対応すべき課題の一つに、国際的な研究基盤や標準規格に関する情報が、データ整備を担う研究者・専門職に十分共有されていないことがある。どの規格に準拠すれば国際的に共有・再利用可能なデータになるのか、どの研究基盤と接続すれば世界の研究資源との相互運用性を確保できるのか、といった基本的情報が行き渡っていない。その結果、日本で作成された研究データは国際標準との整合性を欠き、国際的な知識循環に組み込まれにくい状況が生じている。国際動向に対する適切な情報の収集と発信は、質的データ基盤を構築・発展させる上での重要な前提条件であり、制度的整備・人材育成・国際標準対応をこれらの情報共有と併せて進めることで、初めて日本の研究基盤が国際的に接続可能で持続的なものとなる。

Unicode、TEI、IIIF、DDI などの国際標準規格は、質的研究データを国際的に流通させるための重要な基盤である。しかし、日本ではこれらの標準を研究データ構築の段階で適切に導入し、その改訂や拡張に参画する体制が十分に整っていない。

特に、日本語や東アジア言語の表記体系に固有の課題については、既存の国際標準規格では十分に対応されていない場合も多い。したがって、日本はこれらの標準を単に受容するだけでなく、その形成過程に主体的に参画し、東アジア言語の要件を国際標準に反映させていく必要がある。研究データの長期的利用に不可欠な永続的識別子の運用についても、統一的な指針が十分に確立されていないため、対応が必要である。

(d) 人材育成と教育体制の不十分さ

質的研究データの構築・管理・公開を担う専門人材は、日本において依然として不足している。質的データは、媒体の種類、記述形式、研究文脈、倫理的配慮など、多様な要素を含む複雑な資料である。そのため、その適切な管理には、情報技術の知識

だけでなく、人文学・社会科学の研究方法、資料学的知識、倫理的配慮などを含む複合的な専門性が求められる。

しかし、日本ではこうした専門性を体系的に育成する教育体制が十分に整備されているとは言い難い。DH などの教育プログラムは徐々に拡大しているが、人文学・社会科学分野における多様な研究方法論に十分に対応しきれているわけではなく現時点ではまだ限定的である。さらに、質的研究データの管理や公開を担う専門人材の育成という観点からも、依然として体制は不十分である。

また、大学図書館等において永続性が求められる URI が変更される事例が見られることからも、図書館・博物館・文書館などの専門機関における研修体制の整備が課題となっている。

(e)　持続可能性の欠如

日本の質的研究データ整備は、多くの場合、期間限定の研究プロジェクトや個別機関の活動として進められている。その結果、プロジェクト終了後にデータの維持・公開が困難になる事例も少なくない。

研究成果として蓄積されたデータが社会的資産として長期的に活用されるためには、安定した予算と専門人材を備えた中核的組織が必要である。質的研究データの構築や管理、共有、検索、長期保存には多岐にわたる技術的課題が伴い、それらは技術環境や国際的な規格動向の変化に応じて継続的に更新されるため、その対応には専門性を備えた人材による持続的な取組みが不可欠である。しかし現状では、こうした人材の雇用の多くが任期付きであり、専門知識が任期の満了とともに組織から失われ、継承・蓄積されないという構造的な問題がある。(d)で述べた育成・研修体制の不備が専門人材の供給を妨げる問題であるのに対し、これは育成された人材とその専門性を組織に定着させられない問題であり、両者は別個の原因として作用しつつ、複合して基盤の持続可能性を損なっている。これに加えて、長期的に基盤を支える制度的枠組み自体が確立されていないため、整備されたデータの維持・公開が中断し、研究の透明性や再利用性が損なわれる結果を招いている。

(f)　萌芽的取組みの継続的発展の必要性

2024 年度に開始された文部科学省の「人文学 DX 事業」は、日本における研究データ基盤整備の重要な第一歩である。しかし、現時点では対象とすべき領域に対して事業規模が小さく、テキストデータの構造化と地理情報に関するガイドラインの策定や人材育成のための枠組みの構築といったごく基礎的な取組みにとどまっており、現在はまだ萌芽的段階にあるといえる。これを一過性の事業として終わらせるのではなく、制度的に安定した研究基盤へと発展させていく必要がある。

なお、文部科学省科学技術・学術審議会人文学・社会科学特別委員会「今後の人文学・社会科学の振興に向けた推進方策について(審議のまとめ)」(2025 年 1 月)[57]においても、人文学・社会科学におけるデータ基盤の構築・充実、国際標準化された規格への対応、データ基盤の構築・運用を担う人材の育成等の必要性が指摘されており、本章で述べた課題への認識は政策面でも共有されつつある。本見解では、なかでも特

に質的データに焦点を当て、その固有の課題を踏まえた体制整備の具体像と国際的な規格形成への参画の道筋を提示することを目指すものである。

4　課題に対応した具体策

(1)　制度的に支えられた基盤整備の推進

人文学・社会科学・CSS を横断する「質的研究データ基盤コンソーシアム」を設立し、文部科学省などが基盤的経費として長期的に支える運用可能な制度的枠組みを構築する。欧州の DARIAH や CLARIN が ESFRI ロードマップに位置付けられ、またドイツの NFDI が国家レベルの安定した資金と制度の下で持続的な基盤整備を進めているように、日本でも恒常的なインフラとしての制度化を図らなければならない。具体的には、(2)で述べる中核機関をハブとして、国文学研究資料館、東京大学史料編纂所、立命館 ARC、国立国語研究所、SSJ データアーカイブ、神戸大学 CCSS など既存の拠点を接続し、分野を超えた全国的枠組みとして制度化することが求められる。

(2)　持続可能性を担保する中核機関の設置

中核機関は、新たな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の研究組織を母体として、これに中核機能を持たせる形で設置し、研究データの収集・保存・公開を統括する体制を整備する。母体としては、大学共同利用機関や大学の附置研究所・拠点など、共同利用・共同研究の枠組みを有する組織形態が想定され、法人格は母体組織の形態に従う。いずれの形態においても、安定的な基盤的経費と常勤職員の配置により長期的なガバナンスを担保することが制度的要件となる。

この中核機関は、①質的データの収集・保存・公開を標準化する、②人文学・社会科学・CSS の各分野を橋渡しする、③国際基盤との連携を調整する、という三つの役割を担う必要がある。これにより、日本の研究者が生成したデータを国内外で持続的に利用可能とする。

(3)　国際標準への積極的参画とルール形成

ここでいう国際的なルール形成とは、国家間の交渉を指すのではなく、TEI 協会、IIIF 協会、Unicode 協会、DDI アライアンスといった研究者・研究機関のコミュニティが担う国際標準規格の策定・改訂の過程を指す。東アジアの研究者との連携による地域的な合意形成は、その一部をなすとともに、東アジアの要件を国際標準規格に反映させるための経路でもある。TEI、IIIF、DDI 等の国際標準規格に対する継続的な参画体制を構築し、規格策定・改訂プロセスへの関与を制度化することで、日本語・日本文化資料に固有の一連の課題を検討しつつ国際的に展開するためのスキームを確立する。人文学における TEI や IIIF、社会科学における DDI、文字コードに関する Unicode はいずれも国際的に広く用いられており、日本もその受容者にとどまるのではなく、制度的前提を踏まえつつ、規格整備に主体的に参画すべきである。

例えば TEI では、2021 年に日本語のルビ表記ルールが追加されたが、これは日本の研究者が国際的議論に参加した成果である。このような事例をモデルとしつつ、日本語や東アジア言語の特性を反映させるべく、規格改訂を継続的に国際的な場へ提案する体制

を整える必要がある。また、社会科学分野においても、DDI の枠組みを参照しつつ質的調査データの国際共有を可能にする標準化作業に参加し、日本独自のデータを国際的文脈に位置付けることが重要である。さらに、このことは、単に日本の特性を反映させるというだけでなく、それを通じて、国際標準規格をはじめとする国際的な枠組みや学術的議論の場における多様性拡大への貢献という、大きな学術的・社会的意義を持つ。

(4) 質的データの責任ある共有に向けた体制の整備

データ共有の課題に対応するためには、データの性質、権利関係、来歴、社会的影響に応じた多層的なアクセス制御と、関係する専門家コミュニティ及び当事者との継続的な協議に基づく対応の枠組みを整備する必要がある。具体的には、メタデータの水準では可能な限り発見可能性を確保した上で、データ本体については、完全公開、登録研究者限定、申請制による制限公開、非公開など、段階的なアクセス制御を可能とする仕組みを整える。CESSDA が示すアクセス階層や、SSJ データアーカイブにおける利用申請制度は、その先行事例として参照し得る。

あわせて、いずれの階層においても利用条件をメタデータとして機械可読な形で明示することが再利用の前提となる。公開可能なデータについては、クリエイティブ・コモンズ（CC）ライセンス等の国際的に通用する標準的なライセンスや権利表示を用いることで、利用者が利用可否を個別に問い合わせることなく判断できるようにし、公開に制約を伴うデータについては、その制約の内容と利用申請の手続きを同様に明示する。

また、社会的配慮を要する歴史資料については、研究者コミュニティ、所蔵機関、地域社会、当事者との対話に基づき、公開範囲、利用条件、文脈説明等に関する個別のガイドラインを定める必要がある。植民地支配や戦争等に伴って本来の所在地や帰属共同体のもとから他者の手に渡った資料については、その経緯を含む来歴情報（provenance）を構造化メタデータとして明示し、FAIR 原則に加えて CARE 原則[3]等も視野に入れた協議体制を整える。著作権保護期間が満了した資料については、標準化された権利記述語彙を用いて権利状態と再利用条件を明示し、適切な再利用を促進する。これらを持続的に運用するため、中核機関の下に分野横断的なデータ倫理・データ管理の専門家委員会を設置し、分野別・資料類型別のガイドラインや事例集を継続的に更新する体制を整えることが求められる。

(5) 教育と人材育成の強化

大学院教育において質的研究データ基盤に関する共通科目を設置し、専門人材の体系的育成を行う。欧州の CESSDA や DARIAH では、データ管理や質的データ処理を学ぶための教育リソースやサマースクールが提供され、学生・研究者・専門職の人々が学習機会

[3] CARE 原則（CARE Principles for Indigenous Data Governance）とは、先住民族データ・ガバナンスのための原則である。Global Indigenous Data Alliance（GIDA）が 2019 年に公表した。Collective Benefit（集合的利益）、Authority to Control（管理の権限）、Responsibility（責任）、Ethics（倫理）の四原則からなり、データの背後にある共同体の権利と利益を重視する。データの発見・アクセス・相互運用・再利用というデータ側の特性に着目する FAIR 原則を補完するものとして位置付けられ、先住民族に限らず、不均衡な力関係の下で収集された資料の公開設計にも援用されている。

を得られるようになっている。ドイツの Konsort SWD では、博士課程教育と連携したデータ管理教育が制度的に組み込まれている。

日本でも、東京大学の次世代人文学開発センター、立命館大学の文化情報学専修、同志社大学文化情報学部・研究科、九州大学人文情報学大学院といった既存の拠点を横断的に連携させ、全国的な教育プログラムを形成することが必要である。その際、コンソーシアムに参画する各機関が、それぞれが強みを持つ資料の種類や研究方法論に即した人材育成を担い、全体として人文学・社会科学の多様な質的データに対応できる人材層を形成することが望ましい。加えて、人文学 DX 事業では人材育成ハブを設け、主に研究者向けの教育プログラムを開発しており、これを実効性のあるものとして公表することが期待される。社会科学分野では SSJ データアーカイブと連動した「質的調査データの公開・管理」に関する教育を整備し、CSS 分野では CCSS や SICSS Tokyo を軸に「大規模データ解析やシミュレーションと質的研究データ基盤を接続する教育」を強化することが求められる。

また、図書館・博物館・文書館等の資料に関わる専門職を対象にした研修を制度化し、IIIF や TEI といった国際標準規格を実装レベルで理解できる人材を増やす必要がある。それとともに、教材用のデータセットやツールを容易に使えるように整備しつつ、DARIAH が提供する dariahTeach のように、動画を含むオンライン教材をオープンアクセス化して国内外で共有できる仕組みを作ることも不可欠である。

このようにして育成した人材の適切な評価は、前述の中核機関において検討されるべき重要な事項である。国際的には様々な議論があるが、それらの議論を踏まえて策定されているものとして米国の文学系の学会である MLA(Modern Language Association)による「デジタル学術研究を評価するためのガイドライン」[58]や、アメリカ歴史学協会による「歴史学におけるデジタル研究を評価するためのガイドライン(Guidelines for the Evaluation of Digital Scholarship in History)」[59]等がある。ただし、米国では大学図書館員等の専門職が教員に準じた身分と研究上の役割を担う制度が広く整備されているのに対し[60]、日本では専門職の育成・評価・人事制度の整備は現在も政策課題として検討途上にあり[61]、専門職と研究者の役割分担の前提が異なるためにそのまま適用することはできないが、このテーマを検討する上で参考とすべきである。

(6) 情報共有の強化と国内外ネットワークの構築

国際的研究基盤及び標準規格に関する情報を集約する国内ポータルを整備し、継続的な情報発信と共有を行う。総合知に代表されるような分野横断型研究の実現を目指すためには、DANS、CESSDA、NFDI、DARIAH、CLARIN といった質的研究データに関わる国際的基盤について、研究者や専門職が分野を超えて仕組みを理解できる環境が必要となる。

このような環境を構築するためには、①国際的な質的データ基盤の最新動向を共有するための国内ポータルの設立、②国際標準に関連する規格文書のアップデートに応じた継続的な翻訳と国内への普及、③教育プログラムにおいて、国際的基盤の動向や対応方法に関する講座等を組み込む、④研究者・専門職が参加できる、基盤に関する国内外の共同研究やワークショップの定期的開催、といった具体策を講じる必要がある。国際的

な動向の可視化と共有を制度化することは、日本の研究基盤が国際的に機能するためには必須の条件である。

(7) 対応策の実現によってもたらされる効果

ここまで提示した六つの対応策は相互に独立したものではなく、相補的に機能する関係にある。そのため、これらを一体的に推進することではじめて十分な効果が発揮される。制度的に支えられた基盤整備は、研究資源・成果を長期的に保存し再利用可能にする土台となる。国際標準規格の議論への継続的参画は、日本の研究資源・成果を国際的知識循環に接続し、日本や東アジア固有の文化的特性を世界に反映させる契機となる。ELSI への対応は、以下に詳述するように、研究及び基盤運用に対する社会的信頼を確保するとともに、国際的なルール共有に基づく共同研究の基盤を形成する。教育と人材育成は、研究者及び専門職の層を厚くし、基盤の持続的運用を支える。中核機関の設置は、これらの取組みを制度的・財政的に統合し、個別的・断片的な活動を全国的インフラへと組織化する役割を担う。さらに、情報共有の強化は、研究者や専門職が国内外の動向を把握し基盤運用に主体的に関与することを可能にする点で、これらすべての取組みを支える基盤条件となる。

このように、各対応策は相互に連関しながら機能することにより、個別には達成し得ない統合的効果を生み出し、質的研究データ基盤を持続的かつ発展的に機能させる基礎を形成する。

5 質的研究データ基盤における ELSI への対応

(1) 基本的視点

データ駆動型研究が社会の広範な領域に展開する中で、人文学・社会科学においても ELSI（Ethical, Legal, and Social Issues； 倫理的・法的・社会的な課題）は不可欠な枠組みである。とりわけ DH や CSS は、人の活動と密接に関わる多様なデータを対象とするため、その研究手法及び成果が社会に直接的な影響を及ぼし得る。このため、国際標準規格に関わる事項を含む質的研究データ基盤の整備・運用においては、データ管理の問題にとどまらず、倫理的・法的・社会的観点を統合的に組み込む必要がある。

重要なのは、ELSI を単なるリスク管理や規範遵守の枠組みとしてではなく、基盤構築の前提条件として位置付けることである。すなわち、どのようなデータを収集し、いかに構造化し、どのような形で公開・共有するのかという基盤設計そのものが、倫理的・社会的判断を内包している。

さらに、人文学における批判的視点は、こうした判断の前提を問い直す役割を担う。すなわち、データやモデルが前提とする知識体系や価値観、分類枠組みを検証することは、AI 倫理において求められる透明性や説明責任の基盤ともなる。この意味で、ELSI は単なる規範的枠組みではなく、人文学的批判と AI 倫理を接続する理論的基盤としても位置付けられる。

(2) 倫理的・社会的課題と基盤設計

CSS における行動ログや SNS データの利用に伴うプライバシーや再識別リスク、DH における歴史資料のデジタル化に伴う差別や権力関係の再生産の問題は、いずれもデータの取扱いにとどまらず、基盤の設計と密接に関わる。

データの選択、記述、公開の範囲と方法は、研究対象の可視化と不可視化を規定し、結果として知識の構造そのものに影響を与える。そのため、当事者コミュニティの視点や社会的文脈を踏まえた設計が不可欠であり、ELSI は基盤の設計指針として機能する必要がある。

こうした基盤設計上の課題が集約的に現れるのが、インタビュー記録や逐語録、フィールドノートなど、人を対象とする調査研究を通じて取得されたデータと、調査対象者の同意(インフォームドコンセント)との関係である。取得時に得られた同意は、データ基盤への登載や、生成 AI 等の新たな技術による二次利用までを想定していない場合が多く、これのみを根拠として無限定に共有・利活用を進めることはできない。

他方、医学・生命科学分野を主たる淵源とする事前審査型の倫理審査の枠組みと質的研究の方法論的特性との関係は、国内外で制度設計上の課題として認識されている。日本社会学会の研究指針は、倫理審査委員会の社会学分野における有用性について見解が定まっていないとする一方、質的データを含む調査データのアーカイブ化と二次分析への活用を推奨しており[62]、米国の連邦被験者保護規則（2018 年改正 Common Rule）はオーラルヒストリーや歴史学研究等を、審査対象となる「研究」の定義から除外し[63]、カナダの研究倫理方針 TCPS2 は質的研究の方法論的特性に審査を適合させる独立の章を設けている[64]。こうした状況を踏まえれば、同意の範囲と限界への対応は、個々の研究者や機関の審査のみに委ねるのではなく、基盤の側に組み込む必要がある。第４章(4)で述べた段階的なアクセス制御と当事者との継続的な協議の枠組みはまさにこの要請に応えるものであり、その運用に当たっては、取得時の同意の範囲の確認と、利用条件のメタデータによる明示を組み込むことが求められる。あわせて、今後新たに取得されるデータについては、将来の共有・二次利用の可能性を見据えた同意取得の在り方を、各分野の方法論的特性に即して検討していく必要がある。

(3) 法的・国際的枠組み

著作権や個人情報保護に関する制度への対応は、基盤整備の前提条件であると同時に、研究データの流通と国際連携の可能性を規定する。特に、テキスト・画像マイニングやデータ共有に関する各国制度の差異は、研究基盤の国際接続性に直接的な影響を及ぼす。

したがって、ELSI は国内制度への対応にとどまらず、国際的なルールやガイドラインとの整合性を確保しつつ、研究基盤を設計するための枠組みとして位置付ける必要がある。

(4) 学際研究及び AI との関係

AI 研究の進展に伴い、言語・文化・社会構造に関する知見の重要性は高まっているが、同時に、データの偏在や表象の歪みが技術に組み込まれる危険もある。このような状況において、質的研究データ基盤に ELSI を組み込むことは、単にリスクを回避するためで

はなく、研究の質そのものを高めるために不可欠である。特に、人文学における解釈学的・批判的アプローチは、データの背後にある意味構造や権力関係を可視化し、AI システムが前提とする分類や推論の枠組みを相対化する役割を果たす。こうした人文学・社会科学の知見と AI 研究との協働は、まさに学際研究の実践にほかならない。ELSI は、この協働において各分野が共有すべき問題設定と判断の枠組みを提供する点で、分野間の接合面として機能する。すなわち、質的研究データ基盤に ELSI を組み込むことは、学際的協働の共通基盤を整えることでもある。

したがって、質的データは単なる学習資源としてではなく、その生成過程や文脈とともに扱われる必要がある。人文学・社会科学の知見を組み込むことにより、AI の開発・運用における透明性、説明責任、さらには社会的妥当性を確保する基盤が形成される。

(5) 実装とガバナンス

以上の観点を踏まえると、ELSI は質的研究データ基盤の整備に付随する要素ではなく、その成立を支える中核的要件である。そのため、基盤構築の各段階において ELSI を継続的に検討する仕組みを整備し、専門家コミュニティと社会との対話に基づくガバナンスを確立することが求められる。

このように、基盤構築と ELSI を密接に結び付けて推進することにより、日本における人文学・社会科学の研究基盤は、単なるデータの集積を超え、国際的に通用する質の高い学術基盤へと発展する。これは、研究の透明性や再現性の向上、学際的・国際的共同研究の促進といった、次章で述べる諸効果を実現するための前提条件となる。

6 期待される効果

(1) 研究の透明性と再現性の向上

質的データ基盤の整備は、人文学・社会科学の双方において、研究の透明性と再現性を大きく高める効果をもたらす。人文学では、翻刻や注釈を加えた史資料を共通の規格に沿って構造化し、根拠資料とともに公開することで、研究成果の証拠性を明確に示すことが可能となる。海外でも、既述の Text+のように、テキスト資料を国際標準規格に基づき保存・共有する基盤が学術成果の信頼性を支えており、前述の国書データベース・SHIPS・SAT なども整備の結果として同様の基盤となり得る。

社会科学においても、ELSI に配慮しながらインタビュー記録やフィールドノートなどの質的調査データを適切に匿名化した上で保存し、二次利用を可能にすることは、分析の妥当性や再検証可能性の向上に直結する。国内では、SSJ データアーカイブがそのための仕組みを早期から整備し、研究コミュニティに広く貢献してきた。このような動きが更に広がることが期待できる。

このように、海外・国内の取組みが示す通り、質的データを国際標準規格に基づいて保存し、透明性の高い形で共有することは、研究成果の信頼性を強化し、異なる研究分野間の知の循環を促進する。

(2) 学際的・国際的共同研究の促進

質的データ基盤は学際的かつ国際的な共同研究の推進にも資する。既述のとおり、欧州では CLARIN が言語資源の多国間共有を、CESSDA が量的・質的データの統合利用（混合研究法の基盤）を、DARIAH が歴史資料・文学テキストの国際的共有を、それぞれ学際的・国際的な共同研究の基盤として実現している。日本でも、人文学の構造化テキストデータ、社会科学の質的調査データ、CSS のログやシミュレーションモデルを共通基盤上で接続し、分野横断的に利用できるようにすれば、これまで分断されてきた研究領域を越えた新たな知見が生まれる可能性が更に高まる。例えば、歴史資料と現代社会調査データを組み合わせた比較研究や、文学作品の分析と SNS データの解析を統合した研究など、様々な国際的かつ学際的な研究領域が開かれるだろう。

(3) 教育と人材育成への貢献

教育と人材育成の面でも、質的データ基盤の効果は大きい。人文学では、翻刻や注釈データを教材として利用することで、学生が史資料の読解に加え、デジタル的な表現や構造化の方法を学ぶことができる。既述の dariahTeach に加え、欧州各国の大学が共同制作するオンライン教材 TEI by Example[65]など、研究データ構築のためのオンライン教育資源は既に国際的に広く共有され、研究と教育をつなぐ基盤となっている。社会科学では、DANS や CESSDA が提供するトレーニングを通じて、質的データの匿名化やアーカイブ化の手法が教育の中に組み込まれている。CSS の領域でも、SICSS Tokyo のような取組みを通じて、国際的に共有された教材を用いて質的・量的データを統合的に分析できる人材が育成されている。日本においても、これまで DH を推進してきた大学が有する教育資源を有機的に連携させ、国際的教材を取り込む形で教育体系を形成すれば、人文学・社会科学・CSS を横断できる専門人材を持続的に育成することが可能になる。

(4) 社会への知識共有と政策形成支援

さらに、質的データ基盤は学術界だけでなく社会への知識共有や、政策形成の支援にも直結する。人文学に関わる多様な文化的資料の公開は、生涯学習や文化継承の資源となるとともに、観光や地域振興、日本列島諸文化の国際理解にも寄与する。立命館大学 ARC による浮世絵や映画資料の公開を通じた国際利用促進は、その好例である。社会科学の質的データは、防災や福祉などの政策立案の基盤となる。CSS における量的データも併せたシミュレーションや SNS データの解析は、社会課題の予測やリスク分析に活用され、社会全体の意思決定に資する。人文学の質的データにおいても、歴史的典籍や地域調査データを基盤化することで、市民、教育機関、政策担当者が利用できる知識基盤となり得る。これまでにも歴史的文書に残された古地震をはじめとする災害記録の防災における有益さは確認されており、同様の事例が更に増えることが期待される。

(5) AI 時代における研究基盤としての意義

これらの活動は、AI for Science[66]及び生成 AI を含む AI 研究開発の推進にも大きく貢献する。ELSI を踏まえて適切に作成・共有された機械可読性の高い質的データを継続的に提供することは、各種モデルにとって高品質な学習データの供給源となり、日本語・東アジア言語の特性を踏まえた自然言語処理や、文化的・社会的文脈の理解、社会

調査データに基づく推論の高度化を支える。同時に、こうしたデータを分析支援に活用することで、研究データの自動生成・注釈付与・翻刻支援・分類・要約といった作業が効率化され、研究者はより高度な分析や新規テーマの探索に注力できるようになり、研究手法そのものの革新にもつながる。

さらに、質的データ基盤の整備は、AI 研究の健全な発展にも資する。言語・文化・社会構造への深い理解を要する局面が増える中で人文学・社会科学が適切に関与することは、データ偏在や文化的バイアス、社会的不公正の再生産といったリスクを抑制し、AI 技術の透明性・説明責任や社会受容性の向上に寄与する。加えて、学術的信頼性を備えた質的データは、生成 AI の出力を批判的に検証する参照基盤としても機能し、AI を通じた情報流通の質を高める。

(6) SDGs 及び公共的知識基盤への貢献

質的研究データ基盤は、持続可能な開発目標（SDGs）の達成にも貢献する。公開データを誰もが教材として利用できる環境は教育機会の平等を支え、社会の多様な声や経験を質的研究データとして記録・共有することは不平等の是正に資する。また、透明性の高いデータ基盤は、平和と公正を基盤とする社会の構築にも寄与する。さらに、こうした基盤はデジタル時代における公共的知識インフラとしての性格を持ち、学術研究の成果を社会全体で再利用可能な形で蓄積することで、知識の偏在を是正し、持続可能な知識循環を支える。質的研究データが国際的に接続可能な形で整備されることは、日本の研究成果をグローバルな知識空間に位置付けると同時に、東アジアを含む広域的な知識連携の基盤形成にもつながる。

欧州の事例が示す通り、研究データ基盤は技術的仕組みにとどまらず、社会全体の持続可能性や民主的社会を支える公共的インフラとして機能する。日本が各分野の強みを活かして基盤整備を進めることは、国際社会への貢献であると同時に、デジタル時代にふさわしい「社会に開かれた学術」を実現する重要な一歩となる。

7 今後の展望

(1) ロードマップの構築

これまで述べてきたとおり、日本における質的データ基盤整備のためには制度的な持続可能性が不可欠である。生成 AI の急速な発展を踏まえた短期的なロードマップとともに、人文学・社会科学における知識の持続性を踏まえ、10 年から 20 年の期間を想定した長期的なロードマップをも策定し、段階的に基盤を強化することが必要である。短期的ロードマップとしては、国内 10 機関・10 分野程度の専門的な質的データ構築が可能な研究機関の参加を募り、２年ごとに質的データの構築・集約とそれを踏まえた AI 活用について検討と実践、そして成果公表を行うというサイクルを３回、６年間程度継続することで、AI 時代における人文学・社会科学の質的データの意義を国内外に広く認知させる。この短期的ロードマップは、人文学 DX 事業やデータインフラ事業など既に進行している事業と、参加機関が既に有するデータ構築の体制を土台とし、時限的な事業経費によって実施することを想定している。一方、長期的ロードマップの第一ステップで

は、４章(2)で述べたとおり既存の研究組織に中核機能を持たせる形で中核機関を発足させ、母体となる組織の既存の人員・設備と、既存事業との連携及び競争的資金等の時限的な経費を用いて、既存の取組みの接続・調整と、国際標準に準拠した規格及びその改訂への関与プロセスの確立を進める。第二ステップでは、第一ステップで実績を示した中核機関を基礎に、時限的な事業経費に依存する体制から恒常的な基盤的経費と常勤職員による体制へと転換し、分野横断的な基盤を制度化する。欧州の DARIAH や CLARIN も、ERIC として制度化される以前に、時限的な資金による準備段階を数年間経ており、この二段階の道筋はそれに倣うものである。そして第三ステップでは、海外の研究基盤と連携し、国際的な枠組みの中で中核的な役割を担う段階へと展開する。

(2) 国際連携における日本の責任と可能性

日本は､国際 DH 学会連合の国際会議や TEI 国際会議を東アジア地域で最初に開催し、国際的な研究者コミュニティとの協働を積み重ねてきた経験を有している。この経験は、東アジア各地の研究者コミュニティとの対等な協力の下で、地域の質的研究データ基盤の形成に貢献する上での足がかりとなる。東アジアには、日本語、中国語、韓国語、漢文資料をはじめとする多様な言語文化資源が存在し、それらを国際標準規格に基づいて共有可能な形で整備することは、世界的にも独自性の高い貢献となる。

一方、地域的な連携についてみると、社会科学分野では、日本の SSJ データアーカイブを含む各国・地域のデータアーカイブによるネットワークである NASSDA(Network of Asian Social Science Data Archives)が 2016 年に発足し、データ共有と相互利用に向けた情報交換と定期的な会合を重ねている[67]。ただし、その対象は量的な社会調査データが中心であり、上述の多様な言語文化資源を含む質的データを扱い得る地域的な連携枠組みは確立されていない。この空白を埋める上で、既存の連携の経験は重要な足がかりとなる。

そのため､日本が東アジア諸国､あるいは東南アジア諸国とも連携しつつ TEI、IIIF、DDI、Unicode などの国際標準規格の議論に継続的に参画し、アジアに共通する課題を国際標準規格に反映させる体制を整えることを目指す。これは、日本国内における研究の利便性を高めるにとどまらず、地域全体の知識基盤を国際的な枠組みの中で適切に位置付けることにつながる。この取組みにより、日本は国際的質的データ基盤において単なる利用者ではなく、東アジア各地の研究者コミュニティと連携しつつ、規格形成や制度設計に積極的に貢献する主体へと発展することを目指す。

(3) 学術政策との整合と推進

質的データ基盤の構築と運営は、学術研究支援の枠にとどまらず、学術政策の柱として位置付ける必要がある。「第７期科学技術・イノベーション基本計画に向けての提言」では、第６期と比較しても、質的研究データ基盤が政策形成や社会課題解決に直結する不可欠な要素としてより明確に位置付けられており、特に AI や自然言語処理を活用した質的データ分析の高度化が期待されている。この高度化の実現は、倫理やプライバシーを保護しつつ質的データの共有や再利用を可能とする基盤整備に直結するものであ

る。人文学・社会科学の各分野で得られる知見を統合することにより、複雑化する社会課題に対処するための知識を提供し得る点も重要である。

そのためには、基盤整備に関わる法的枠組みの整備、恒常的な予算措置の確保、そして対話を継続的に行う回路の整備が求められる。対話の場としては、総合科学技術・イノベーション会議、文部科学省科学技術・学術審議会の関連する委員会、日本学術会議など、既に制度化された場が存在する。求められるのは新たな会議体の設置ではなく、４章で述べたコンソーシアム及び中核機関が、人文学・社会科学の質的データ基盤を代表する恒常的な主体としてこれらの場に参加し、基盤の整備状況と課題を継続的に伝える回路を確立することである。さらに、国内における基盤モデル開発が政策課題となっている現在、日本語・東アジア言語の高品質な学習・評価データの供給源としての質的データ基盤の意義は、AI 政策の文脈にも接続され得るものである。このため、AI 研究開発を担う機関や関連する政策部門との協議も、この回路の一部として位置付けるべきである。これらの施策を段階的かつ着実に進めることで、質的データ基盤を中核とした持続可能な学術政策の実現と、それに整合した研究の推進へとつながる。

(4)　社会に開かれ、社会とともに歩む学術基盤へ

質的データ基盤は、研究者のみの資産ではなく、社会全体が参照し、参加し、共に育てていくべき公共的な知識基盤である必要があることは述べてきたとおりである。人文学においては、信頼度の高い質的データ基盤を通じて、市民が歴史や文化にアクセスし再解釈できる環境を整えることで、文化継承や近年注目を集めつつある Public Humanities（※）を実践し得る基盤となる。社会科学においても、調査データを透明性の高い形で基盤化し、市民・研究者・政策当局がアクセス可能な形で整備することは、エビデンスに基づく社会の意思決定を支える土台となる。さらに、シミュレーションや大規模データ解析を社会課題の理解や予測に活かし、市民とともに社会の将来像を構想するための学術的支援を可能にする点も重要である。

このように、質的データ基盤が公共的な資源として展開されることで、学術と社会は相互補完的に発展し、市民・研究者・政策当局による協働を通じて知識を創出し共有する循環が可能となる。このような基盤を整備することによって、「社会に開かれた学術」を更に発展させ、「社会とともに歩む学術」となり、学術が社会全体の未来構築に主体的に関わる形を実現していくことになる。

８　見解のまとめ

質的データ基盤の整備は、人文学・社会科学・CSS に共通する重要課題であり、学術研究の高度化のみならず、社会の持続可能性を支えるために必要な活動である。翻刻・注釈・インタビュー記録・フィールドノート・シミュレーションモデルなど多様な質的データを、オープンサイエンスを踏まえ ELSI に配慮しつつ、国際標準規格に準拠し、かつその形成・改訂に参画しながら保存・共有することで、研究の透明性と再現性は大きく向上し、学術的知見の確かな蓄積が可能となる。

こうした基盤を社会に開くことは、日本の人文学・社会科学の存在意義の再認識にもつながる。文化資料や調査データの公開は、生涯学習・文化継承やエビデンスに基づく政策形成を支え、信頼性の高い質的データは生成 AI の学習資源及びその出力を検証する基盤として、AI 技術の健全な発展にも不可欠である。国際的にも、日本は東アジア各地の研究者コミュニティと連携しつつ国際標準規格の形成に参画し、日本語・漢字文化圏の特性を国際的知識基盤に反映させる役割を担い得る。

この未来像を実現するために、人文学・社会科学を横断し長期的に運営される「質的研究データ基盤コンソーシアム」及び持続可能性を担保する中核機関を設置し、国際標準への積極的参画とルール形成、教育と人材育成の強化、そして情報共有の強化と国内外ネットワークの構築を一体的に進める必要がある。これらを通じて、質的データ基盤を社会全体の公共的資源として確立していくことが求められる（図１）。

本見解の実現に向けては、関係する主体がそれぞれの立場から役割を担うことが期待される。文部科学省には、質的研究データ基盤を学術政策上に位置付け、コンソーシアム及び中核機関を支える恒常的な制度と予算の枠組みを整えること（４章(1)(2)、７章(3)）、及び大学院教育における質的研究データ基盤に関する教育の展開を支援すること（４章(5)）が期待される。日本学術振興会には、データインフラ事業を通じて蓄積された成果と機関間連携を、本見解の構想する基盤へと発展的に引き継ぐとともに、競争的資金を通じて質的データの構築と人材育成を支えることが期待される。国立国会図書館には、デジタルコレクション及びジャパンサーチを通じた大規模なデジタル資料の提供と、永続的識別子や権利表示など研究データの長期利用を支える技術的基盤の整備、並びに図書館等の専門職の研修における主導的役割が期待される（４章(4)(5)）。人間文化研究機構には、大学共同利用機関法人として、人文学における質的データの構築・共有の実績と共同利用の枠組みを活かし、コンソーシアムの形成と中核機能の確立に主体的に関与することが期待される（４章(2)）。これらに加え、各大学・研究機関、学協会及び研究者コミュニティが、データの構築、国際標準規格への参画、教育の実施の担い手として基盤を支えることが不可欠である。

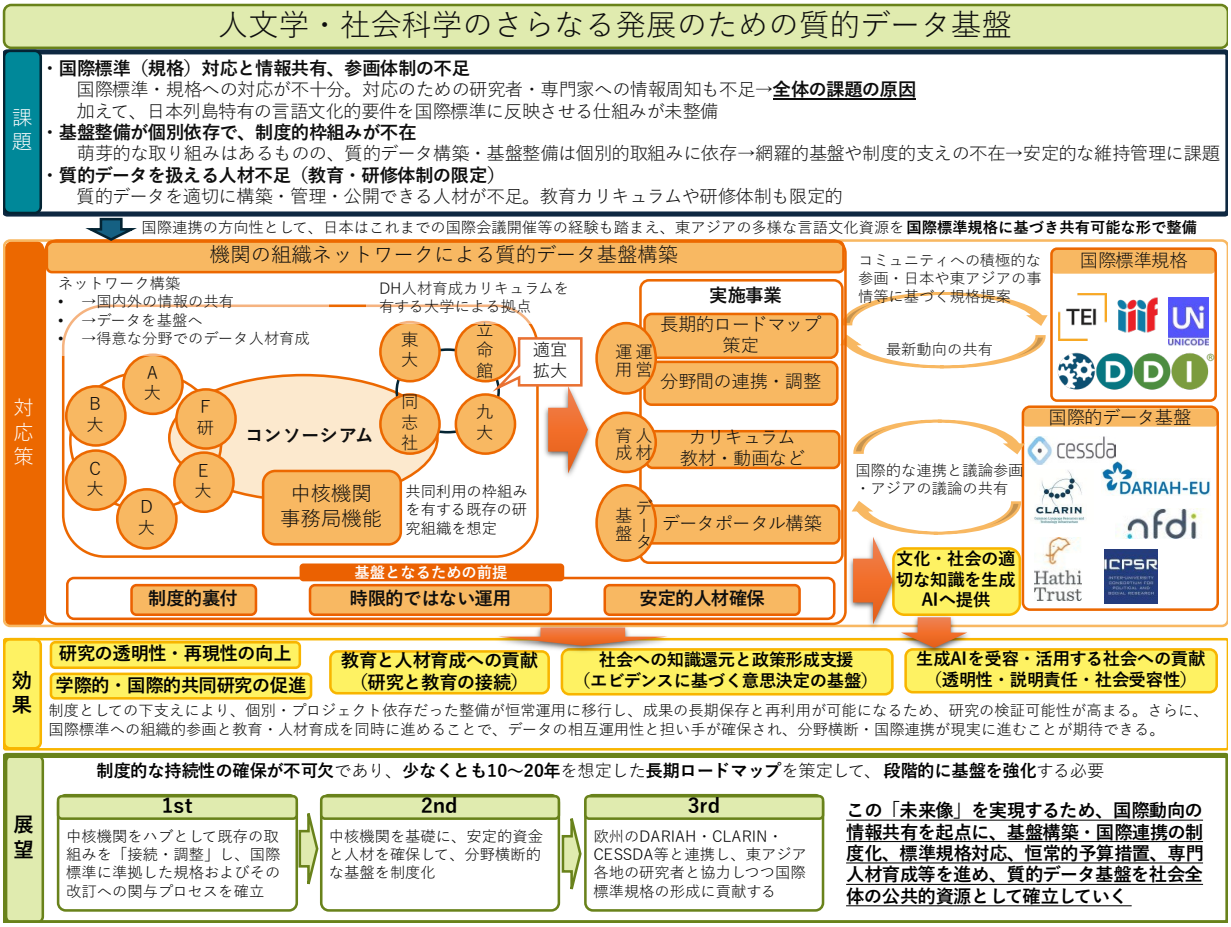


図１　オープンサイエンスを踏まえた質的研究データ基盤の全体像

（出典）本分科会作成

＜用語説明＞

① CLARIN

CLARIN（Common Language Resources and Technology Infrastructure）は、言語資源と言語処理技術を対象とする欧州規模の研究インフラであり、言語学のみならず人文学・社会科学全体におけるデータ駆動型研究を支えることを目的として構築されてきた。2000 年代半ば、各国で整備されていたコーパスや辞書、音声資料、言語解析ツールが相互運用されていないという課題を背景に、これらを横断的に接続する共通基盤として構想された。

CLARIN は 2012 年に欧州研究インフラ・コンソーシアム（ERIC）として正式に設立され、現在は CLARIN として多数の加盟国による分散型インフラとして運営されている。その特徴は、各国が運営するセンター（CLARIN Centres）を認証制度によって結び付け、標準化されたメタデータ、永続的識別子、認証・認可基盤を通じて、言語資源への安定的かつ横断的なアクセスを実現している点にある。

また、CLARIN はツール提供にとどまらず、研究データ管理、法的・倫理的課題への対応、教育・トレーニングの整備にも力を注いでおり、言語資源を「再利用可能な研究データ」として流通させる制度的枠組みを形成してきた。ドイツでは CLARIN-D が DARIAH-DE や TextGrid と連携し、言語資源とテキスト研究基盤とを接続する役割を担っている。

このように CLARIN は、言語データを核としつつ、人文学研究全体を支える欧州的研究インフラの一角として、DARIAH や Huma-Num と補完的な関係を築きながら発展してきたのである。

② DARIAH

DARIAH（Digital Research Infrastructure for the Arts and Humanities）は、ヨーロッパにおける人文学・芸術研究のための分散型研究インフラであり、デジタル時代の研究実践を支えることを目的として構築されてきた。2000 年代後半、各国で進展していた DH プロジェクトや研究基盤が国境ごとに分断されているという課題を背景に、これらを横断的に結び付ける欧州規模の枠組みとして構想された。

DARIAH は 2014 年に欧州研究インフラ・コンソーシアム（ERIC）の正式メンバーとなり、以後 DARIAH として運営されている。その特徴は、単一の巨大システムを構築するのではなく、各国が保有するツール、リポジトリ、サービス、人材ネットワークを相互接続する「分散型インフラ」という点にある。これにより、学術編集、テキスト分析、研究データ管理、教育・研修といった多様な活動を柔軟に支援している。

ドイツでは、TextGrid や CLARIN-D など既存の人文学基盤が DARIAH-DE を通じて DARIAH に統合され、国内インフラと欧州レベルの研究環境とを接続する役割を果たしてきた。すなわち DARIAH は、個別プロジェクトや国別基盤を超えて、人文学研究を持続的かつ国際的に展開するための制度的・技術的な枠組みとして位置付けられている。

③ HathiTrust 研究センター

HathiTrust 研究センター（HathiTrust Research Center, HTRC）は、北米と中心とした大規模デジタル図書館である HathiTrust の所蔵データを対象に、計算機的手法を用いた研究を可能にするために設立された研究基盤である。HathiTrust 自体は 2008 年に米国の大学図書館を中心とする共同事業として発足し、Google Books 等を通じてデジタル化された膨大な書籍を長期保存・提供することを目的としてきた。

HTRC はこの流れを受け、2011 年にインディアナ大学とイリノイ大学アーバナ＝シャンペーン校を中心に設立された。その目的は、著作権の制約下にある資料を含む大規模コレクションに対しても、研究者がテキスト分析や統計的手法を適用できる「安全な研究環境」を提供する点にある。全文を直接閲覧できない資料であっても、特徴量抽出や非消費型（non-consumptive）研究という枠組みを通じて、学術研究への利用を可能にしている。

HTRC では、トピックモデリングや語彙分析、ジャンル分析などのためのツールや API（アプリケーション・プログラミング・インターフェイス）が整備され、図書館主導のデジタル化資源を計算人文学へと接続する役割を果たしてきた。このようにHathiTrust Research Center は、デジタル図書館とデータ駆動型人文学研究を結び付ける米国を代表する研究インフラとして位置付けられている。

④ Huma-Num

Huma-Num は、フランスにおける人文学・社会科学（SHS）のための中核的な研究インフラであり、デジタル時代の学術研究を支える共通基盤として整備されてきた。2000 年代以降、研究データのデジタル化とオンライン公開が急速に進む一方で、個別プロジェクトごとにデータ管理や公開方法が分断されているという課題が顕在化し、その解決を目的として設立された。

Huma-Num は 2013 年にフランスの「非常に大規模な研究インフラ（TGIR）」として位置付けられ、研究データの長期保存、永続的識別子の付与、標準に基づく公開支援などを担っている。その役割は、単にデータを保管することにとどまらず、TEI や IIIF などの国際標準を前提としたデータ整備を促し、学術情報流通の質を高める点にある。

また、Huma-Num はフランス国内の研究機関やプロジェクトを束ねるハブとして機能すると同時に、欧州レベルでは DARIAH などの国際的研究インフラとも連携している。このようにして、Huma-Num は研究者主導の人文学データを持続的に支える制度的・技術的基盤として、フランスの DH を支えてきた。

⑤ Public Humanities（公共人文学）

Public Humanities（公共人文学）とは、人文学の研究を大学や専門家の内部にとどめず、市民・地域社会・図書館・博物館・文書館・メディア等と協働しながら、社会とともに実践する人文学の在り方を指す。研究成果を一方向的に社会へ発信するアウトリーチにとどまらず、市民が資料の翻刻・注釈や歴史・文化の解釈に参加し、研究者と市民が知識を共同で生み出すこと（co-creation）を重視する点に特徴がある。2010 年代以降、英語圏を

中心に大学院プログラムや専門の学術誌が設立されるなど、人文学の一領域として定着しつつある。

デジタル環境においては、公開された資料や構造化データを市民が自由に利用・再解釈できることと、市民参加型の翻刻・注釈プロジェクトを通じて市民がデータ構築そのものに参加できることが、その実践の基盤となる。日本においても、市民参加によって歴史資料の翻刻を進める「みんなで翻刻」のような取組みが、この方向の実践例として位置付けられる。本見解が 7 章(4)で述べる、社会全体が参照し参加する公共的な知識基盤としての質的データ基盤は、こうした Public Humanities の実践を支える土台となるものである。

⑥ TEI（Text Encoding Initiative）ガイドライン

TEI ガイドライン（Text Encoding Initiative Guidelines）は、人文学研究で扱われるテキスト資料を、コンピュータで利用しやすい形に構造化して記述するための国際的な指針である。その策定は 1987 年に始まり、学術テキストを分野横断的に記述する共通仕様を目指して国際的な研究者コミュニティである TEI 協会によって進められてきた。最初の本格的なガイドラインは 1994 年に公開されている。

対象は古典文学、歴史文書、写本、碑文など多岐にわたり、単なる文字列としてではなく、章・段落・注記・人物名・地名・異文といった構造や意味的役割を明示的に記述することを可能にする。

TEI は XML という記述言語を採用し、共通の規則に基づくことで検索や分析、長期保存、研究成果の共有を容易にする。また、研究目的や資料の性質に応じて要素を選択・調整できる柔軟性を備えている点も重要である。このように、TEI Guidelines は、DH における基盤的な標準として現在まで継続的に改訂・発展してきた。

TEI ガイドラインにおける東アジア・日本語対応をめぐっては、2016 年に東アジア／日本語分科会（SIG: Special Interest Group）が設立されたことが重要な転機となっている。この分科会は、漢字文化圏に特有の文字体系や文献形式が、従来の欧米語中心の想定では十分に扱いきれないという問題意識を共有する研究者によって組織されたものである。同分科会の意義は、単に「日本語向けの拡張」を提案することにとどまらず、漢字・仮名・訓読・返り点・ルビ・異体字といった東アジア固有の記述課題を、TEI 全体の設計思想の中でどのように位置付けるかを理論的・実践的に検討する点にある。これにより、ローカルな慣行を場当たり的に追加するのではなく、国際標準としての一貫性を保ちながら、多様な文献文化を包摂する方向性が明確化された。

また、この分科会は、日本や東アジア地域で蓄積されてきた実践例や教育経験を共有し、ガイドライン本文や例示の改善、既存要素の再解釈、新たな要件の整理などを通じて、TEI Guidelines の持続的な改訂に寄与してきた。すなわち、2016 年の分科会設立は、TEI を真に多言語・多文化的な学術基盤へと発展させるための制度的枠組みを与えた点に、その役割と意義があるといえる。

⑦ TextGrid

TextGrid は、ドイツにおいて構築された代表的な人文学研究インフラであり、デジタル学術編集やテキスト研究を支援するための統合的な環境として発展してきた。2000 年代半ば、デジタル化された一次資料や学術編集版を、個別プロジェクトごとに分断された形ではなく、共通基盤の上で作成・保存・公開する必要性が強く認識されるようになり、その流れの中で TextGrid 構想が立ち上げられた。

TextGrid は 2006 年頃からドイツ連邦教育研究省（BMBF）の e-Science 構想（D-Grid）支援を受けて本格的に開発が進められ、テキストの編集・注記・管理を行うための TextGrid Laboratory と、成果を長期保存・公開するための TextGrid Repository から成るインフラとして整備された。とりわけ、TEI に基づく学術編集版の作成を前提とし、複数研究者による共同編集や、長期的なデータ保存を視野に入れた設計が特徴である。

その後、TextGrid はドイツの研究データ基盤を統合する動きの中で、DARIAH-DE や NFDI（National Research Data Infrastructure）といった枠組みと連携し、人文学研究データの持続的な管理と流通を担う役割を担ってきた。このように TextGrid は、個別プロジェクト支援から国家的研究データ基盤への接続へと段階的に位置付けを変えながら、ドイツにおける DH インフラの中核を形成してきたのである。

⑧ UK Data Service

UK Data Service は、英国研究・イノベーション機構（UKRI）を構成する経済社会研究会議（ESRC）の資金提供により運営される、経済・人口・社会科学分野の研究データの中核的基盤である。2012 年に発足し、エセックス大学 UK Data Archive を中心として、マンチェスター大学、Jisc、ユニヴァーシティ・カレッジ・ロンドン（UCL）、エディンバラ大学との共同体制をとり、データの収集・保存・キュレーション・提供に加え、研究者への利用支援や研修、研究データ管理に関する情報提供を担っている。質的データについては、インタビューの逐語録、日記、フィールドノート、自由記述、音声・映像等の非数値的資料を対象とし、その質的二次データのコレクションは世界最大規模とされ、二次利用に必要となる調査文脈の把握や、将来の利用に対する同意を含む法的・倫理的問題等の論点整理も行っている。こうした質的データの受入・キュレーション・長期保存を実務面で支えるのが中核機関 UK Data Archive であり、同機関は 1994 年に設置された質的データ専門の国立サービス Qualidata を統合し、質的データ用メタデータ標準 QuDEx や、検索閲覧システム QualiBank における段落単位の永続識別子・引用方式の開発等、質的データ共有の技術基盤の整備を主導してきた。国際的には、UK Data Archive を通じて欧州の社会科学データアーカイブ連合 CESSDA（2017 年より欧州研究基盤コンソーシアム＝ERIC）の活動に創設時から参画し、国際標準の形成に寄与している。

⑨ 学術編集版

学術編集版（scholarly edition）とは、写本や版本などの一次資料に基づき、研究者が明示的な方法論と根拠を示しながら構成した学術的テキストである。そこでは、本文の確

定だけでなく、資料の来歴、編集方針、注記、異文の扱いなどが体系的に提示され、読者が編集上の判断過程を追跡できることが重視される。critical edition（校訂テキスト／批判版）は、scholarly editionの一形態であり、特に本文批評と異文の整理に重点を置いた編集形態を指す。

近年では、こうした学術編集の枠組みをデジタル環境で実装したデジタル学術編集版（digital scholarly edition）が重要な研究形態として位置付けられている。デジタル学術編集版では、本文、注記、異文、底本情報、さらには原資料画像などを相互に関連付けて提示でき、単一の本文を提示するだけでなく、複数の読みや解釈の可能性を構造的に示すことが可能となる。

また、編集方針や判断の前提をデータとして保持・公開できる点に特徴があり、再利用性や検証可能性を高める研究基盤としての意義を持つ。このように、デジタル学術編集版は、従来のscholarly editionの学術的厳密さを継承しつつ、その表現力と可塑性を拡張するものと位置付けられる。

⑩　持続可能性

本見解において研究データ基盤やデータについて用いる場合、基盤が長期にわたり安定的に維持・運用され得ることを指す。これは、欧州のESFRI(欧州研究インフラ戦略フォーラム)などにおいて研究基盤の長期的な「sustainability (long-term sustainability of research infrastructures)」として国際的に定着している概念に対応するものであり、財政・人材・制度・技術の各面で基盤を維持し得る状態を含意する。なお、第６章(6)等で「社会の持続可能性」というときは、SDGs(持続可能な開発目標)の文脈における意味で用いる。

⑪　デジタル・ヒューマニティーズ

デジタル・ヒューマニティーズ(DH)は、Susan Schreibman, Ray Siemens, John Unsworth編『A Companion to Digital Humanities』（2004, Blackwell Publishing）の刊行を契機に人文学において情報技術を応用することに関するあらゆるテーマを扱う学問領域を指す言葉として用いられるようになった。なお、デジタル人文学、人文情報学、文化情報学という呼称もほぼ同じものを指す言葉として用いられている。

人文学には史学・文学・哲学のみならず、言語学・社会学・心理学・人類学・地理学等の様々な分野が含まれており、研究手法も研究対象も多様である。DHは、情報技術を活用して個々の研究の発展を目指すのみならず、情報技術を媒介とすることで細分化された多様な分野間の対話の機会を提供し、それを以て人文学の発展を志向するものでもある。

欧米においては DH は図書館や図書館情報学との協力関係の下に発展してきた。有力大学のDHセンターは図書館に設置されることが多く、DH司書やデータ司書等としてDHで博士号を取得した研究者を雇用する例もよく見られる。そのようなことから、DH分野の国際学会でも図書館関係者の発表は多く見られる。

DH を多様な分野の協働の場として成立させるための理念的背景として、「方法論の共有地（Methodological Commons）」という概念がよく持ち出される。方法論の共有地とは、個々の分野における様々な手法論にデジタル技術を適用する際に生じる共通項を議論する場を指す。DH では、学会のみならず、個々の研究プロジェクトの中や様々な種類の会合や SNS 等のデジタル媒体を介した場においても、異分野間での対話に積極的に取り組む必要があり、研究活動が方法論の共有地の実践の場となっている。

＜参考文献＞

[1] 日本学術会議科学者委員会学術体制分科会、提言「第７期科学技術・イノベーション基本計画に向けての提言」、2024 年 11 月 28 日.

[2] 日本学術会議社会学委員会社会統計調査アーカイヴ分科会、提言「社会調査をめぐる環境変化と問題解決に向けて」、2017 年９月 19 日.

[3] 日本学術会議社会学委員会東日本大震災後の社会的モニタリングと復興の課題検討分科会、提言「社会的モニタリングとアーカイブ―復興過程の検証と再帰的ガバナンス―」、2020 年９月 14 日.

[4] 日本学術会議オープンサイエンスの深化と推進に関する検討委員会、提言「オープンサイエンスの深化と推進に向けて」、2020 年５月 28 日.
https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-24-t291-1.pdf

[5] 日本学術会議オープンサイエンスを推進するデータ基盤とその利活用に関する検討委員会、回答「研究ＤＸの推進－特にオープンサイエンス、データ利活用推進の視点からー に関する審議について」、2022 年 12 月 23 日.
https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-25-k335.pdf

[6] 日本学術会議第三部理工系学協会の活動と学術情報に関する分科会、提言「学術情報流通の大変革時代に向けた学術情報環境の再構築と国際競争力強化」、2020 年９月 28 日. https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-24-t297-6.pdf

[7] 日本学術会議第一部人文・社会科学の役割とその振興に関する分科会、提言「学術の総合的発展をめざして―人文・社会科学からの提言―」2017 年６月１日.
https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-t242-2.pdf
https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-t248-7.pdf

[8] 日本学術会議社会学委員会Ｗｅｂ調査の課題に関する検討分科会、提言「Web 調査の有効な学術的活用を目指して」、2020 年７月 10 日.
https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-24-t292-3.pdf
https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-24-t296-5.pdf

[9] 日本学術会議総合工学委員会科学的知見の創出に資する可視化分科会、報告「科学的知見の創出に資する可視化 ―文理融合研究と新パラダイム策定―」、2020 年９月８日. https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-24-h200908-2.pdf

[10] 日本学術会議言語・文学委員会・哲学委員会・史学委員会・地域研究委員会合同アジア研究・対アジア関係に関する分科会、提言「新たな情報化時代の人文学的アジア研究に向けて――対外発信の促進と持続可能な研究者養成――」、2017 年９月 21 日
https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-t247-10.pdf

[11] 日本学術会議 情報学委員会 提言「生成 AI を受容・活用する社会の実現に向けて」、2025 年２月 27 日. https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-26-t381.pdf

[12] European Research Infrastructure Consortium(ERIC)(公式 Web サイト)
https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-

innovation/our-digital-future/european-research-infrastructures/eric_en
[13] Consortium of European Social Science Data Archives (CESSDA) (公式 Web サイト) https://www.cessda.eu/
[14] Nationale Forschungsdateninfrastruktur(公式 Web サイト) https://www.nfdi.de/
[15] CERN and OpenAIRE, Zenodo https://zenodo.org/
[16] Ubiquity Press, Journal of Open Humanities Data https://openhumanitiesdata.metajnl.com/
[17] International institute of social history, Research Data Journal for the Humanities and Social Sciences https://researchdatajournal.org/
[18] Social Sciences and Humanities Open Cloud, Social Sciences & Humanities Open Marketplace, https://marketplace.sshopencloud.eu/
[19] Digital Humanities @ Oxford, Digital Humanities at Oxford Summer School https://digital.humanities.ox.ac.uk/digital-humanities-oxford-summer-school
[20] DARIAH, dariahTeach https://teach.dariah.eu/
[21] Institute for Social Research, ICPSR (公式 Web サイト) https://www.icpsr.umich.edu/sites/icpsr/
[22] Digital Humanities Summer Institute(DHSI)(公式 Web サイト) https://dhsi.org/
[23] International Image Interoperability Framework(IIIF)(公式 Web サイト) https://iiif.io/
[24] DDI Alliance, Data Documentation Initiative https://ddialliance.org/
[25] Text Creation Partnership(TCP)(公式 Web サイト) https://textcreationpartnership.org/
[26] Faculty of Linguistics, Philology and Phonetics, University of Oxford, British National Corpus http://www.natcorp.ox.ac.uk/
[27] Distant Reading for European Literary History (COST Action CA16204) https://www.distant-reading.net/
[28] Fischer, Frank, et al. (2019). Programmable Corpora: Introducing, an Infrastructure for the Research on European Drama. In *Proceedings of DH2019: "Complexities"*, https://zenodo.org/records/4284002; DraCor (公式 Web サイト) https://dracor.org/
[29] 国立国語研究所、現代日本語書き言葉均衡コーパス https://clrd.ninjal.ac.jp/bccwj/
[30] 国立国語研究所、国語研日本語ウェブコーパス https://masayu-a.github.io/NWJC/
[31] 東京大学史料編纂所、ships-データベース検索 https://wwwap.hi.u-tokyo.ac.jp/ships/
[32] SAT 大正新脩大藏經テキストデータベース研究会 (公式 Web サイト) https://21dzk.l.u-tokyo.ac.jp/SAT/

[33] 国文学研究資料館、国書データベース https://kokusho.nijl.ac.jp/
[34]立命館大学アート・リサーチセンター（公式 Web サイト）
https://www.arc.ritsumei.ac.jp/
[35]国立国会図書館次世代システム開発研究室（公式 Web サイト）
https://lab.ndl.go.jp/service/tsugidigi/
[36] 永崎研宣、「CA1989 - 動向レビュー：IIIF の概要と主要 API バージョン 3.0 の公開」『カレントアウェアネス』No.346、国立国会図書館、2020 年 12 月 20 日
https://current.ndl.go.jp/ca1989
[37] 日本デジタル・ヒューマニティーズ学会、「デジタル・ヒューマニティーズ」、J-STAGE https://www.jstage.jst.go.jp/browse/jadh/-char/ja
[38] 人間文化研究機構 DH 推進室、DH コンソーシアムプロジェクト、
https://dh.nihu.jp/dihuco
[39] 神戸大学 計算社会科学研究センター（公式 Web サイト） https://ccss.kobe-u.ac.jp/
[40] Journal of Computational Social Science, Springer
https://link.springer.com/journal/42001
[41] Summer Institute in Computational Social Science(SCISS)（公式 Web サイト）
https://sicss.io/
[42] Summer Institute in Computational Social Science 、SICSS Tokyo
https://sicss.io/2022/tokyo/
[43] 東北大学大学院文学研究科・文学部 計算人文社会学研究室（公式 Web サイト）
https://www.sal.tohoku.ac.jp/jp/research/specializations/lab/---id-37.html
[44] 東京大学大学院人文社会系研究科・文学部附属次世代人文学開発センター人文情報学部門（公式 Web サイト） https://dh.l.u-tokyo.ac.jp/
[45] 立命館大学文学研究科文化情報学専修、専修概要
https://www.ritsumei.ac.jp/gslt/major/major13.html/
[46] 同志社大学文化情報学部（公式 Web サイト） https://www.cis.doshisha.ac.jp/
[47] 大阪大学人文学研究科、人文学林 https://jinbungakurin.hmt.osaka-u.ac.jp/
[48] 文部科学省「デジタルと掛けるダブルメジャー大学院教育構築事業 ～X プログラム～」 https://www.mext.go.jp/a_menu/koutou/kaikaku/x-pro/index.html
[49] 九州大学人文情報連係学府（公式 Web サイト） https://dh.kyushu-u.ac.jp/
[50] 千葉大学卓越大学院プログラム、「アジアユーラシア・グローバルリーダー養成のための臨床人文学教育プログラム」 https://jinbun-takuetsu.chiba-u.jp/top/
[51] 人間文化研究機構、「令和７年度　社会との共創によるデジタル・ヒューマニティーズ（DH）の推進　DH ツール構築研究課題の募集について」
https://dh.nihu.jp/news/post/20250425
[52] 一般財団法人人文情報学研究所、TEI 古典籍ビューワ
https://tei.dhii.jp/teiviewer4eaj

[53] 一般財団法人人文情報学研究所、「人文学のためのテキストデータ構築入門」フォローアップサイト　https://www.dhii.jp/dh/tei/
[54] The North American Coordinating Council on Japanese Library Resources（公式 Web サイト）　https://guides.nccjapan.org/homepage
[55] European Association of Japanese Resource Specialists(EAJRS)(公式 Web サイト）https://www.eajrs.net/
[56] 部落解放同盟中央本部、「古地図・古絵図刊行および展示に対する基本的考え方について」、2003 年 10 月 20 日. http://www.bll.gr.jp/archive/siryo-syutyo2003/guide-seimei-20031110.html
[57] 文部科学省科学技術・学術審議会学術分科会人文学・社会科学特別委員会、「今後の人文学・社会科学の振興に向けた推進方策について（審議のまとめ)」、2025 年 1 月 17 日 https://www.mext.go.jp/content/20250117-mxt_sinkou01-000040144_01.pdf
[58] MLA Committee on Information Technology. Guidelines for Evaluating Digital Scholarship. September 2024. https://www.mla.org/content/download/191995/file/Guidelines-Evaluating-Digital-Scholarship.pdf
[59] アメリカ歴史学協会、「歴史学におけるデジタル研究を評価するためのガイドライン」、菊池信彦・小風尚樹・師茂樹・後藤真・永崎研宣訳、2015 年 6 月. https://www.jadh.org/guidelines-for-the-evaluation-of-digital-scholarship-in-history
[60] Association of College and Research Libraries, ACRL Standards for Faculty Status for Academic Librarians, 2007 年承認、2021 年改訂 https://www.ala.org/acrl/standards/standardsfaculty
[61] 「2030 デジタル・ライブラリー」推進に関する検討会「新しい『デジタル・ライブラリー』の実現に向けて検討すべき課題」2024 年 7 月 1 日 https://www.mext.go.jp/content/20240701-mxt_jyohoka01-000036744_3.pdf.pdf
[62] 日本社会学会「日本社会学会倫理綱領にもとづく研究指針」2006 年 10 月 28 日制定、2023 年３月改訂 https://jss-sociology.org/about/researchpolicy/
[63] U.S. Department of Health and Human Services et al., Federal Policy for the Protection of Human Subjects ("Common Rule"), 45 CFR §46.102(l)(1), 2018 年改正(2019 年１月施行) https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-A/part-46
[64] Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council, Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans – TCPS 2 (2022), Chapter 10: Qualitative Research https://ethics.gc.ca/eng/tcps2-eptc2_2022_chapter10-chapitre10.html
[65] TEI by Example project　https://www.teibyexample.org/

[66] 文部科学省科学技術・学術審議会情報委員会、「AI for Science の推進に向けた基本的な考え方について」、第 44 回資料、2025 年 10 月 6 日
https://www.mext.go.jp/content/20251006-mxt_jyohoka01-000045188_04.pdf
[67] NASSDA (Network of Asian Social Science Data Archives) https://nassda.org/

＜参考資料１＞審議経過

令和６年

６月 25 日　地域研究委員会・言語・文学委員会・哲学委員会・心理学・教育学委員会・社会学委員会・史学委員会・法学委員会・経営学委員会・情報学委員会合同デジタル時代における新しい人文・社会科学に関する分科会（第１回）
役員の選出、今後の進め方について

８月 18 日　地域研究委員会・言語・文学委員会・哲学委員会・心理学・教育学委員会・社会学委員会・史学委員会・法学委員会・経営学委員会・情報学委員会合同デジタル時代における新しい人文・社会科学に関する分科会（第２回）
意思の表出に関する方針について

令和７年

３月 30 日　地域研究委員会・言語・文学委員会・哲学委員会・心理学・教育学委員会・社会学委員会・史学委員会・法学委員会・経営学委員会・情報学委員会合同デジタル時代における新しい人文・社会科学に関する分科会（第３回）
意思の表出の骨子案について

11 月 26 日　地域研究委員会・言語・文学委員会・哲学委員会・心理学・教育学委員会・社会学委員会・史学委員会・法学委員会・経営学委員会・情報学委員会合同デジタル時代における新しい人文・社会科学に関する分科会(第４回)
「見解案」について

＜参考資料２＞シンポジウム開催

第 25 期になる 2022 年１月 22 日、シンポジウム「総合知創出に向けた人文・社会科学のデジタル研究基盤構築の現在」をオンラインで開催し、広く市民との対話の場を設けた。

開催趣旨：
第６期科学技術・イノベーション基本計画において人文・社会科学に期待される総合知の実現には、研究データの構築と利活用が一つの鍵となっており、研究データインフラの構築も徐々に進みつつある。
しかしながら、日本におけるこれまでの人文・社会科学における研究データの構築は、ごく一部の例外を除き、決して十分に実施できてきたとはいえない状況である。この背景には日本ならではの独自の事情や技術的制約など、様々な事情があるが、そうした状況が徐々に改善されてきたこともあり、ようやく人文・社会科学における研究データの本来的な課題に向き合う環境が整いつつある。
本シンポジウムでは、こうした状況を踏まえ、現状の課題を共有するとともに、今後の展開やその応用可能性について検討する。

プログラム：
13:30　開会挨拶　永崎研宣（日本学術会議連携会員、一般財団法人人文情報学研究所主席研究員）
13:40　第一部 人文・社会科学における研究データの構築のための国際標準と課題
・学術研究のための多言語・多文字への対応と多漢字・変体仮名
　－ 高田智和（国立国語研究所准教授）
・人文学向け国際標準に基づく東アジア・日本のテキスト構造化
　－ 永崎研宣（日本学術会議連携会員、一般財団法人人文情報学研究所主席研究員）
・画像公開方式 IIIF と歴史 GIS によるデータ統合と総合知
　－ 北本朝展（ROIS-DS 人文学オープンデータ共同利用センター センター長・国立情報学研究所教授）
15:20　第二部 様々なコンテクストによるデータの構築から総合知へ
・データ駆動型研究方法論の開拓を視野に入れた『人事興信録』研究
　－ 増田知子（名古屋大学大学院法学研究科教授）、佐野智也（名古屋大学大学院法学研究科特任講師）
・クラウドソーシングによるデータ構築から古地震研究へ
　－ 加納靖之（東京大学地震研究所・地震火山史料連携研究機構准教授）
・国立国会図書館のデジタル化事業とそこから産み出されるデータの可能性
　－ 大場利康（国立国会図書館電子情報部長）

17:00　第三部　コメント：総合知に向けた人文社会科学データの観点から

- 矢野桂司（日本学術会議第一部会員、立命館大学文学部教授）
- 橋本隆子（日本学術会議連携会員、千葉商科大学副学長）
- 西田眞也（日本学術会議第一部会員、京都大学大学院情報学研究科教授）

18:00 総合ディスカッション
18:30 閉会

主催：
日本学術会議 心理学・教育学委員会、言語・文学委員会、哲学委員会、社会学委員会、史学委員会、地域研究委員会、情報学委員会合同
「デジタル時代における新しい人文・社会科学に関する分科会」

後援：
アート・ドキュメンテーション学会、情報処理学会人文科学とコンピュータ研究会、情報知識学会、デジタルアーカイブ学会、日本デジタル・ヒューマニティーズ学会、日本図書館情報学会、三田図書館・情報学会