

第14章

オープンな 学術情報流通を 支える テキストデータ 構造化

永崎研宣

version 1.0

2026.3.21 作成

1. 学術情報はどのように流通するのか

1-1. 学術情報流通の変化と課題

学術研究の成果は、論文や書籍として公表されることで社会に共有されてきた。しかし近年、研究成果の流通のあり方は大きく変化している。オンラインでの公開が進み、オープンアクセスやオープンサイエンスといった考え方が広がる中で、学術情報は「公表されているかどうか」だけでなく、「どのような形で流通しているか」が問われるようになってきている。

1-2. 「閲覧できる」ことと「流通する」ことの違い

学術情報が流通するためには、単に閲覧できる状態にあるだけでは十分ではない。検索によって発見できること、他の研究と結びつけて参照できること、必要に応じて再利用できることが重要になる。たとえば、PDF として公開された論文は人間が読むことはできるが、内容の構造や意味関係は多くの場合暗黙のままであり、機械的な処理や横断的な再利用には限界がある。このような形式では、研究成果は「存在している」ものの、「流通している」とは言いにくい。

1-3. 日本における構造化実践：J-STAGE と JATS/XML

日本におけるその代表例が、J-STAGE における JATS/XML の採用である。J-STAGE は、Japan Science and Technology Agency (JST) が運営する、日本の学協会などによる学術論文・会議録・研究報告などの電子ジャーナルを公開・流通させるための総合プラットフォームである。1999 年の運用開始から 25 年以上が経過し、2024 年 5 月時点で 4,000 を超えるジャーナルや会議録等が登録され、約 5,600,000 件以上の論文・記事が収録されている（刊行終了誌や予稿集等を含む）。これらの多くはオープンアクセスで提供され、研究者のみならず誰でも自由に閲覧できる環境が整えられている。

JATS (Journal Article Tag Suite) は、この J-STAGE において論文データを内部管理・流通させるために用いられている XML ベースの国際標準仕様である。JATS は、学術論文の構造を機械可読な形で記述することを目的として策定されており、論文の章立て、抄録、本文の段落、引用文献、図表、著者情報、所属、キーワード、出版履歴などを共通の語彙と構造で明示的に表現できる。これは、米国の標準化団体である National Information Standards Organization (NISO) によって策定・維持され、国際的な学術出版インフラとして広く利用されている。

1-4. 構造化されたテキストが支える学術情報流通

J-STAGE では、論文を PDF として閲覧可能にするだけでなく、内部的に JATS/XML 形式の構造化データを保持・管理することで、検索性や長期保存、他の学術サービスとの連携を支えている。たとえば、論文メタデータのエクスポート、外部データベースとの相互参照、検索エンジ

ンによる索引生成といった機能は、JATS による構造化情報を基盤として実現されている。すなわち、学術情報を単に「存在させる」だけでなく、流通可能な形に組織化し、他サービスや研究者の利活用とつなげていく仕組みが現実に機能しているのである。

このような取り組みが示しているのは、学術情報の流通において、構造化されたテキストが不可欠であるという点である。テキストに明示的な構造が与えられていれば、その内容は検索や集計、可視化、他のデータとの連携といった形で利用されやすくなる。すなわち、構造化されたテキストは、単に読むための情報ではなく、流通し、結びつき、再利用される知識資源として機能し得る。

1-5. TEI による人文学テキスト構造化の位置づけ

TEI ガイドラインに基づくテキストデータの記述も、このような学術情報流通を支える形式の一つである。ただし、TEI は JATS のように主として「論文という成果物」を対象とする規格ではなく、人文学研究における多様なテキスト資料——文学作品、史料、書簡、日記、写本など——をどのように構造化し、どの単位で意味づけるかについて、長年にわたって議論されてきた共通の枠組みを提供している。そのため、TEI によって記述されたテキストは、特定の研究プロジェクトの内部に閉じたものではなく、他の研究や別の文脈と接続される可能性を持つ。

この点で、テキストデータの構造化は、研究者個人の作業効率を高めるための工夫にとどまらない。それは、研究成果を学術情報流通の中に位置づけ、他者にとって利用可能な形に整えるための実践でもある。個々の研究が積み重なることで、学術情報は孤立した成果の集合ではなく、相互に結びついた知識のネットワークとして機能するようになる。

本章では、このような視点から、テキストデータの構造化がどのようにして学術情報流通を支え、再利用や異分野連携、発見性の向上につながっていくのかを、具体的に考えていく。次節では、まず「再利用される研究をつくる」という観点から、構造化されたテキストデータの意義を掘り下げる。

1-6. teiHeader からのリポジトリ登録

TEI データが学術情報流通の対象として扱いやすい大きな理由の一つは、本文 (<text>) とは別に、<teiHeader> が「引用・目録化・処理」のためのメタデータを体系的に保持できる点にある。TEI ガイドライン自身も、teiHeader を、利用者が適切に引用するための書誌情報や、図書館・アーカイブが目録レコードを作成するための記述を与える場として位置づけている。このことは、TEI データをデータリポジトリに載せる際に、本文の構造を無理に別形式へ変換しなくても、ヘッダ部分のメタデータをリポジトリが受け取れるスキーマ（例：Dublin Core や DataCite など）に“クロスウォーク”するだけで、検索・収集・引用に適した形に整えやすいことを意味する。たとえば DataCite は研究データ等を「引用と検索 (retrieval)」のために一貫して同定することを目的としたメタデータスキーマであり、Dublin Core へのマッピングも整備

されているため、teiHeader を起点にした変換方針を設計しやすい。

さらにこの枠組みは、TEI 本来の「本文中心のテキスト」だけでなく、RDF 等のグラフデータにも応用できる。すなわち、研究上の主たるデータが RDF トリプルや関係記述であっても、それを TEI の枠内にシリアルライズし（たとえば <standOff> に関係記述や外部参照をまとめ、本文とは分離して保持するなど）、その上で teiHeader を付与しておけば、当該データは「人文学研究の記述対象」から「リポジトリで流通し、検索され、引用される研究オブジェクト」へと位置づけ直しやすくなる。TEI は本文の階層構造と、standOff による外在化された記述（参照される情報の集約）を併存させる設計を持つため、階層とグラフの双方を視野に入れたデータ編成が可能である。

このような「teiHeader を核にした流通設計」を現実に支えるために、TEI コミュニティは変換や入出力のためのツール提供も進めてきた。たとえば TEI の変換環境（OxGarage など）は、TEI を中心に据えつつ、他形式との相互変換を可能にする入口として広く参照されている。また近年は、TEI と RDF の相互変換や RDF へのシリアルライズといった論点も、コミュニティ内で継続的に扱われている。したがって TEI は、単に「人文学テキストを精密に記述する」ための枠組みにとどまらず、研究成果を（メタデータ変換を含む運用設計によって）学術情報流通のインフラへ接続するための“パッケージ”としても理解できる。これに関しては、本書のサポートサイトを通じて、取組みを進めていきたい。

つまり、teiHeader を核にした流通設計は、研究成果を「公開した」だけの状態から、「検索され、引用され、他の研究と結びつけられる」状態へ移行させるための入口となり得るのである。再利用は、その入口を通してはじめて現実の実践として立ち上がることになる。次節では、この観点から、再利用される研究データとは何かを整理し、TEI によって判断と前提をどのように残すべきかを具体的に検討する。

2. 再利用される研究データに向けて

2-1. 再利用可能性としての研究成果

学術情報が流通するためには、研究成果が単に公開されているだけでなく、**他者によって再利用される可能性を持っていることが重要である**。ここで言う再利用とは、同じ研究をそのまま繰り返すことや、結果を単純に引用することを指すのではない。むしろ、別の研究目的や問いのもとで、同じデータや記述が新たな文脈に置き直され、異なる意味や価値を与えられることを含んでいる。

研究成果が再利用されるかどうかは、研究の規模や完成度そのものによって決まるわけではない。重要なのは、**どのような形で記録され、どの程度まで判断の前提が明示されているか**という点である。前提や選択が暗黙のまま残されている場合、成果はその場限りのものとして消費され

やすく、別の文脈で用いられることは難しくなる。

2-2. 判断と前提を明示する記述としての TEI

この点において、TEI による記述では、とくに <encodingDesc> やその内部に記述される <editorialDecl> が重要な役割を果たす。<encodingDesc> は、そのテキストがどのような方針や前提のもとで符号化されたのかを説明するための要素であり、<editorialDecl> は、正規化や訂正、省略、表記の扱いなど、編集上の判断を明示的に記述する場である。ここに何が書かれているかによって、テキストデータがどのような性格を持つのか、どの範囲まで信頼できるのか、どのような利用が想定されているのかが、他者にとって理解可能になる。

再利用の可能性は、必ずしも同じ分野の研究に限られない。文学作品のテキストが歴史研究や文化研究の資料として参照されることもあれば、歴史資料が地理情報や社会分析と結びつけられることもある。また、構造化されたテキストデータは、検索や可視化、計量的分析、教育用途、さらには生成 AI を用いた分析や支援といった文脈でも利用されうる。このように、再利用とは単一の利用形態を想定するものではなく、**複数の将来像に開かれていることを意味している**。

このとき、再利用される研究をつくるということは、将来の利用者に向けて「正しい解釈」を残すことではない。むしろ、**どのような判断を行い、どのような選択肢の中から現在の記述が成り立っているのかを追跡可能な形で残すことに意味がある**。その判断がどこに記述されているのか、すなわち <editorialDecl> や関連する記述を通じて確認できることが、再利用の前提となる。

2-3. 将来の利用者にかかれた構造化テキスト

TEI による構造化されたテキストは、内容を固定化するのではなく、判断の所在を明確にする。どの部分が本文で、どの部分が注記であり、どの情報が内部リストとして管理されているのかが分かれば、データを受け取る側は、その構造と前提を踏まえて利用方法を検討できる。これは、再利用を単に「可能にする」だけでなく、**利用者に新たな問いや使い方を促す設計**であるとも言える。

このように考えると、再利用される研究をつくることは、研究成果を将来の研究者や利用者に向けて開く行為である。必ずしも大規模で完成されたデータセットである必要はなく、判断を構造として残し、他者がそこから出発できる足場を用意することが重要である。とくに <encodingDesc> や <editorialDecl> によって編集方針や判断の前提が明示されていることは、テキストデータを長期的に再利用可能な研究資源とするための中核的な条件である。

次節では、このような構造化がどのようにして異なる分野や研究領域をつなぎ、分野横断的な対話を可能にするのかについて考える。

3. 異分野をつなぐための最小限の共通構造

3-1. 分野差を越えるために必要な「構造」の共有

研究成果が再利用され、学術情報として流通していくためには、同一分野の内部で完結するだけでなく、異なる分野や方法論と接続される可能性を持つことが重要である。しかし実際には、分野を越えた研究の連携は容易ではない。その大きな理由の一つは、研究対象や問いそのもの以上に、記述の仕方や前提となる構造が分野ごとに異なっている点にある。

たとえば文学研究では、作品の全体構成や章・段落の関係、語りの形式、注記や異文の扱いといった点が重要な関心となる。これは TEI では <text>、<body>、<div>、<p> といった階層構造や、<note>、<app>、<rdg> などの要素によって記述することができる。一方、歴史学においては、人物・出来事・年代・場所といった要素の関係性や時間的推移が重視される。その場合、<persName>、<placeName>、<date>、<event> といった要素や、@when、@corresp、@ref といった属性を用いて、テキスト中の記述を明示的に構造化することが有効となる。

3-2. 固有表現と関係をめぐるグラフ的理解

さらに重要なのは、固有表現とその関係を、個々の出現箇所ではなく、全体として捉える視点である。TEI では、本文中に現れる人名や地名を <persName> や <placeName> として記述するだけでなく、<listPerson> や <listPlace> といった内部リストを用いて、それらを一意の実体として整理し、@corresp や @ref によって本文中の記述と結びつけることができる。このような記述は、テキストを単なる線形の文字列としてではなく、固有表現をノードとし、それらの参照関係をエッジとするグラフ構造として理解することを可能にする。

このグラフ的な視点は、分野横断的な接続において特に重要である。文学研究においては、登場人物同士の関係や作品内での配置が問題となり、歴史学では、人物・場所・出来事の関係ネットワークが分析対象となる。情報学の立場からは、こうした関係構造をデータとして抽出し、検索や可視化、計量的分析に用いることが可能になる。また、近年の AI 研究においても、テキスト中の固有表現とその関係をグラフ構造として扱う手法が広く用いられており、意味的な関係を持つデータであることが、分析や推論の前提となりつつある。

TEI による記述は、このようなグラフ構造を直接的に強制するものではないが、グラフとして解釈可能な形で関係を明示するための記述基盤を提供している。本文中の参照関係、内部リストによる実体の一元化、さらには <standOff> を用いた関係記述などを通じて、階層構造（文書構造）と関係構造（意味的ネットワーク）を併存させることができる点は、TEI の大きな特徴である。

このように、TEI は分野ごとの解釈や方法を統一しようとするものではない。TEI は、内容や結論を揃えるための規格ではなく、どのような実体が存在し、それらがどのような関係にあると判断されたのかを外在化し、共有するための記述枠である。文学研究者は文学研究の関心に基づ

いて、歴史研究者は歴史研究の関心に基づいて、それぞれ異なる重点で TEI を用いることができる。その一方で、固有表現・参照関係・内部リストといった基本的な仕組みが共有されているため、異分野の研究者が互いのデータを理解し、再利用し、別の問いのもとで読み替えることが可能になる。

3-3. TEI を共通言語とする分野横断的接続

この意味で、TEI は異分野を橋渡しするための翻訳装置であり、**階層構造とグラフ構造を併せ持つ共通言語**である。研究成果を TEI によって構造化することは、特定分野の内部にとどまらず、他分野との接続可能性をあらかじめ組み込んだ形で研究を設計することに他ならない。

次節では、このような共通構造と関係記述が、研究成果の発見性をどのように高めるのか、すなわち「見つけられる研究」をどのように設計できるのかについて考える。

4. 発見される研究を設計する

4-1. 発見されない研究という問題

研究成果がどれほど丁寧に作られていたとしても、それが他者に見つけられなければ、学術的な議論の中で参照される機会は限られてしまう。ここで言う「発見」とは、偶然目に入ることを指すのではなく、検索や参照、他の研究との関連付けを通じて、**必要な人に必要な形で到達**することを意味している。

従来の学術研究では、論文誌や書籍という媒体を通じて成果が流通してきた。しかし、デジタル環境においては、研究成果がどのような形式で記録され、どのような情報が明示されているかによって、発見されやすさは大きく左右される。全文検索に頼るだけでは、研究の内容や構造を適切に把握することは難しく、関連する研究同士が結びつかないまま孤立してしまうことも少なくない。

4-2. 構造化がもたらす検索と結びつき

テキストデータが構造化されていれば、発見性は大きく向上する。たとえば、本文中で人名や地名が単なる文字列として書かれている場合と、`<persName>` や `<placeName>` として明示的に区別され、内部リスト (`<listPerson>` や `<listPlace>`) によって同一性が管理されている場合とでは、検索や集計の可能性が大きく異なる。

このような構造化が行われていれば、「同じ人物に言及している研究」や「同じ場所を扱っている資料」を横断的に検索することが可能になる。また、`<date>` 要素や `@when` 属性によって年代が明示されていれば、時間軸に沿った検索や可視化も行いやすくなる。出来事や行為が `<event>` やそれに準ずる形で記述されていれば、出来事を単位とした探索も可能となる。

TEI による記述は、このような検索や結びつきを支えるための **共通言語**として機能する。TEI は、特定の検索システムや可視化ツールに依存する形式ではなく、テキストの意味的構造を明示的に記述するための枠組みである。そのため、異なるシステムやプロジェクトであっても、同じ構造を前提としてデータを扱うことができ、検索結果や分析結果を相互に比較・統合しやすくなる。

4-3. 典拠・外部システムとの接続による発見性の拡張

発見性を高めるうえで重要なのは、単に検索語に引っかかることだけではない。研究成果が、より広い知識空間の中でどの位置にあるのかが分かることも重要である。この点で、人名典拠や地名典拠と接続されたデータは、特定の研究プロジェクトの内部に閉じることなく、他のデータセットや研究成果と結びつく可能性を持つ。

TEI では、@ref や @corresp、@sameAs といった属性を用いることで、本文中の固有表現を外部の典拠情報や識別子と結びつけることができる。これにより、TEI で記述されたテキストは、内部構造を持つだけでなく、外部の知識基盤と接続されたノードとして機能できるようになる。

このような接続が行われていれば、研究成果は単独で存在するのではなく、外部のデータベース、リポジトリ、検索サービス、さらには他分野の研究成果と結びついたネットワークの一部として発見される。これは、研究の主題や対象が、別の分野や別の問いの文脈で再発見される可能性を高めることにつながる。

4-4. AI 時代における発見性と構造化

近年、AI を用いた検索や分析が急速に発展する中で、研究成果の発見性は新たな局面を迎えている。大規模言語モデルや検索支援システムは、大量のテキストを処理することができるが、その理解の精度は、入力されるデータの構造や明示性に大きく依存する。

構造化されていないテキストは、AI にとっても解釈の手がかりが乏しく、文脈や関係性を誤って理解する可能性が高い。一方、TEI によって意味的な単位や関係が明示されたテキストは、AI にとっても扱いやすいデータとなる。人名・地名・出来事・役割といった情報が明確に区別され、参照関係が構造として記述されていれば、要約、検索補助、関係抽出、知識グラフ的な処理の精度も向上しやすくなる。

ここで重要なのは、TEI による構造化が AI のために行われるのではないという点である。あくまで人文学研究における判断を明示的に記述した結果として構造が与えられ、その構造が結果的に AI にとっても有用な情報源となる。この点において、TEI は人間と機械の双方にとって理解可能な共通言語として機能している。

4-5. 発見性を設計するという視点

以上のように、研究成果の発見性は偶然に委ねられるものではなく、**設計することができる**。

どの情報を要素として切り出し、どの関係を属性や参照として明示するかによって、研究成果がどのように見つけられ、どのような文脈で再利用されるかが変わってくる。

TEI を用いたテキスト構造化は、研究成果を未来の探索空間に配置するための実践である。発見される研究を設計することは、研究を一度きりの成果として終わらせるのではなく、他者の間いや技術、さらには新たな分析手法と出会い続ける可能性を持たせることである。その意味で、発見性を意識した構造化は、研究実践を長期的な知識循環の中に位置づけるための重要な視点である。

次節では、このような実践を現実的に支えるために、研究者自身と研究を取り巻く運営側がそれぞれどのような点を意識すべきかについて整理する。

5. 研究実践を支える役割と環境

前節までで見てきたように、テキストデータの構造化は、研究成果の再利用、異分野との接続、発見性の向上を通じて、オープンな学術情報流通を支える基盤となる。しかし、こうした実践は自然に成立するものではなく、研究者個人の取り組みと、それを支える環境の双方が必要となる。本節では、そのために何が求められるのかを、研究者側と運営側の視点から整理する。

5-1. 研究者側に必要なこと

研究者に求められるのは、最初から完成度の高いデータを作ることではない。むしろ重要なのは、判断を行ったこと自体を記録し、後から理解できる形で残すという姿勢である。どの単位を構造として切り出したのか、どの表記を同一の実体として扱ったのか、なぜその判断を採用したのかといった点を、できる範囲で明示しておくことが、再利用や議論の前提となる。

また、構造化は一度で完成させる必要はない。研究の進行に応じて記述を見直し、判断を修正したり、補足したりすることは自然な過程である。TEI を用いた記述は、そのような試行錯誤を前提とした枠組みであり、**研究ノートを構造化して残す**という感覚で取り組むことができる。

さらに、将来の利用者を漠然とでも想定することが有効である。それは特定の誰かである必要はなく、「数年後の自分」や「同じ資料を扱うかもしれない他の研究者」で十分である。そのような視点を持つことで、内部リストの作成や典拠との接続、記述方針の明示といった作業が、単なる手間ではなく、研究の射程を広げるための投資として位置づけられるようになる。

5-2. 運営側に必要なこと

テキストデータの構造化を通じた学術情報流通を持続的なものとするためには、研究者個人の努力だけでなく、それを受け止め、蓄積し、共有するための運営的な基盤が不可欠である。大学や研究機関、学会、研究プロジェクトといった運営側には、構造化された研究成果を保存・公開

できるリポジトリの整備、記述や公開に関する相談に応じるサポート体制、教育やワークショップを通じた基礎的なリテラシーの共有などが求められる。

ここで重要なのは、「運営側」と「研究者側」を固定的に分けて考える必要はないという点である。多くの場合、研究者は同時に、研究会の世話役やプロジェクトのメンバー、データ管理の担当者、教育の担い手といった形で、運営側の役割を担うことになる。また、自身の研究をよりよい形で残し、次につなげたいと考えるならば、**敢えて運営側に回る**という選択肢も現実的である。

近年、国際的には、こうした研究と運営、記述と利活用、理論と実装の間を担う専門的な役割として、**Research Software Engineer (RSE)** という概念が確立しつつある。RSE は、研究者と協働しながら、研究データやソフトウェア、ツール、システムを設計・開発・維持する役割を担う存在であり、単なる技術支援ではなく、研究の前提や方法論を理解したうえで実装に関与する点に特徴がある。国際的には、RSE を専門職として位置づけ、コミュニティや学会、ネットワークを形成する動きも進んでいる。

TEI を用いた研究実践においても、こうした RSE 的な役割は重要な意味を持つ。構造化されたテキストデータを利活用するための検索・閲覧・可視化・分析ツールや、TEI データを他のデータ資源と接続するシステムの開発は、単なる補助的作業ではない。それらは、TEI による記述の設計思想や判断の前提を踏まえて行われるものであり、その過程で生じる設計上の選択や工夫もまた、研究実践の一部として位置づけることができる。

この意味で、データの記述・公開、ツール開発、システム運用は、分離された営みではなく、**相互に支え合うエコシステム**を構成している。TEI による研究では、テキストをどのように記述するだけでなく、その記述をどのように利用し、提示し、他者に届けるかという点まで含めて実践が広がっていく。そのような実践に関わること自体が、研究コミュニティの知識基盤を形づくる行為であり、研究の一環として十分に意味を持つ。

このように、運営側の役割や RSE 的な実践は、研究者から切り離された特別な立場ではなく、研究実践の延長線上にあるものとして捉えることができる。構造化されたテキストデータを受け止める場や仕組み、そしてそれを活用するツールやシステムを整えることは、他者の研究を支えると同時に、自身の研究が将来どのように参照され、使われていくのかを左右する行為でもある。

その意味で、テキストデータの構造化をめぐる運営やツール開発、RSE 的な役割は、誰か一部の専門家に任せるべき負担ではなく、研究コミュニティ全体で分担し、引き受けていくべき実践である。研究者が判断を記録し、運営側がそれを受け止め、RSE 的な実装によって利活用を可能にするという関係が、相互に行き来し得る形で設計されることで、はじめて持続的な学術情報流通が成立する。

6. おわりに — 学術情報流通と生成 AI 時代を支える実践としてのテキスト構造化

本書で扱ってきたテキストデータの構造化は、個々の研究を進めるための技法であると同時に、研究成果を他者と共有し、将来に引き継ぐための実践でもある。人名や地名の同一性を明示し、判断の前提を構造として記述し、再利用や検証が可能な形でデータを提示することは、研究成果をオープンな学術情報流通の中に位置づける行為である。

研究はしばしば、個別の論文や一つの成果物として完結するものとして捉えられがちである。しかし、テキストデータを構造化し、記述方針や判断を明示的に残しておくことで、その研究は他者にとって参照可能な資源となり、別の問いや別の研究目的のもとで再利用される可能性を持つようになる。このことは、研究成果の寿命を延ばし、研究の射程を広げることにもつながる。

また、こうした実践が積み重なることで、学術情報は個々の研究プロジェクトの内部にとどまらず、検索や再利用、異分野連携を通じて流通可能な知識資源となる。その意味で、テキストデータの構造化は、研究者個人の作業であると同時に、学術全体の知識基盤を下支えする営みの一部でもある。

さらに近年では、生成 AI をはじめとするデジタル技術が、学術情報の利用や探索のあり方を大きく変えつつある。生成 AI は大量のテキストを処理し、要約や検索支援、関係抽出などを行うことができるが、その振る舞いは、入力されるデータの質や構造に強く依存している。構造化されていないテキストは、AI にとって文脈や関係性を把握しにくく、誤った解釈や不適切な一般化を生む可能性も高い。

これに対して、TEI に基づいて構造化されたテキストデータは、人文学研究における判断や意味の単位が明示されているため、生成 AI にとっても扱いやすい知識資源となる。重要なのは、AI のために構造化を行うのではなく、**人間の研究判断を明示的に記述した結果としての構造が、結果的に AI の利用可能性を高める**という点である。このようなデータは、生成 AI による分析や支援を、人文学研究の文脈に即した形で行うための、信頼性の高い基盤となりうる。

この意味で、テキストデータの構造化は、オープンな学術情報流通を支えるだけでなく、生成 AI 時代において、学術的知識がどのように利用され、再解釈されていくのかを方向づける実践でもある。構造化された研究成果が蓄積されることで、生成 AI による学術支援や知識探索も、より透明で説明可能な形で進められるようになるだろう。

本書で行ってきた演習は、TEI を操作できるようになること自体を目的とするものではない。むしろ、研究判断をどのように構造として記述し、共有し、他者の理解や再利用、さらには生成 AI による利活用にも耐える形で提示するかを考えるための基礎を身につけることを目的としている。本書が、読者自身の研究を学術情報流通の中に位置づけ、次の研究や実践、そして生成 AI 時代における学術と社会をつなぐ知識基盤の形成へとつなげていくための出発点となれば幸いである。