

第13章

TEI 実践を 研究として 位置づける

永崎研宣

version 1.0

2026.3.21 作成

1. 研究としての位置づけ

1-1. 記述の過程に含まれる解釈・判断

本書で扱ってきた TEI ガイドラインに基づくテキストの構造化や記述は、実務的には多くの時間と労力を要する作業である。しかし同時に、その成果がどのような位置づけを与えられるのかは、必ずしも自明ではない。TEI による記述は、研究の準備作業や分析に入る前の下処理として扱われることもあれば、単なる技術的支援やデータ整備として研究成果の外に置かれてしまうこともある。

では、ここでは、なぜこうした作業を研究として捉えることをあえて検討しようとしているのだろうか。その理由は、TEI による記述の過程で行われている判断が、人文学研究においてしばしば重視され得る解釈行為と密接に関わっているからである。テキストをどの単位で区切るのか、どの表記を同一の実体として扱うのか、どの情報を本文に残し、どの情報を別の層として管理するのかといった選択は、いずれも資料をどのように読解し、どのような前提で理解するかという解釈に基づく判断である。

もしこれらの判断を「研究以前の作業」として切り離してしまえば、研究の前提となる解釈や選択が不可視化され、最終的な結論だけが提示されることになる。その結果、研究がどのような前提のもとで成り立っているのかを、他者が十分に理解することが難しくなる場合もある。

1-2. 可視化がもたらす議論可能性

一方で、研究の理解や評価においては、結論そのものに加えて、その結論に至るまでにどのような前提と判断が置かれてきたのかを明示できるかどうかの問題となることがある。TEI による記述は、こうした判断を構造として外在化し、可視化するための手段を提供する。判断が可視化されることによって、他者はその前提を理解し、同意することも、批判することも、別の解釈を提示することも可能になる。

この点において、TEI による記述を研究として検討することの意義は、作業の価値を一方的に主張することにあるのではない。むしろそれは、研究を相互に議論可能なものとして捉え直すための視点を提供することにある。判断が可視化され、共有されることによって、研究は個人の内部で完結する営みではなく、他者との対話の中で洗練されていく過程として理解されるようになる。

1-3. 協働的な理解への接続

このような相互の検討や再考の積み重ねは、学術研究において重要な営みの一つである。TEI による記述は、そのための共通の参照点を与え、研究者同士が同じ地平に立って議論することを可能にする。結果として、個々の研究は孤立した成果にとどまらず、協働的により高い理解水準

を目指す営みの一部として位置づけられる余地を持つことになる。

このような問題意識を踏まえ、本章では、TEI ガイドラインに基づくテキストの構造化や記述を、単なる「準備作業」や「分析の前段階」としてではなく、研究の前提や判断を可視化する実践として捉え直す。

2. TEI 記述はどの段階で「研究」になるのか

2-1. 成果の単位としての TEI データ

以上を踏まえると、次に問題となるのは、TEI による記述がどの段階で研究成果として評価され得るのかという点である。研究においては、「何が成果として評価されるのか」という問いが常につきまとう。従来の人文学では、研究成果の主要な単位は論文や学位論文であったが、TEI に関する成果が公表されることの多いデジタル・ヒューマニティーズ (DH) 分野においては、成果の単位は必ずしもそれに限られない。

TEI に基づいて作成されたテキストデータには、どの資料を対象とし、どのような構造判断を行い、どの解釈を採用し、どの可能性を退けたのかといった研究上の判断が、データ構造として固定化されている。したがって、TEI データは単に論文の素材となるものではなく、研究判断そのものを体現した成果物として位置づけることもできる。

このとき、研究成果として評価されうるかどうかを左右するのは、そのデータが他者にとってどの程度利用可能であるかという点である。第三者がそのデータを検証できるか、同じ前提に立って再分析できるか、判断の根拠を追跡できるかといった条件が満たされている必要がある。これらの条件を備えた TEI データは、それを適切に説明する論文を伴うことによって、研究成果の一単位として十分に評価され得る。

2-2. 研究評価において問われる説明責任

研究評価の場において、TEI を用いた研究で問われるのは、どれほど多くのタグを付与したかという量的な側面ではない。むしろ、研究目的と構造設計との関係がどれだけ明確に説明されているかが重要である。なぜその単位を要素として切り出したのか、なぜその属性や参照関係を採用したのかが、研究上の問いと結びついて説明されていなければならない。

また、人名や地名の同一性判断、内部リストと外部典拠の使い分け、XInclude によるリスト管理といった設計が、場当たりの判断の積み重ねではなく、一貫した方針のもとで行われているかどうかとも評価の対象となる。これらの判断が明示されていれば、他者はその前提を理解したうえで、同意や批判、別の解釈を提示することが可能になる。

さらに、成果がどのような形で提示され、共有されているかも重要である。TEI データが論文の付随資料としてのみ存在するのか、それともデータ自体が引用可能で、再利用を前提として公開されているのかによって、研究の射程は大きく異なる。後者の場合、研究は個人の成果にとど

まらず、協働的な知的営みの一部として位置づけられる。

以上のように、TEI を用いた記述が研究として評価されるかどうかは、個々の結論や解釈の新規性だけでなく、どのような研究判断が行われ、それがどの程度明示的に記述されているかという点に大きく依存している。次節では、このような評価のあり方を理解するための背景として、オープンサイエンスや FAIR 原則の動向を取り上げる。

3. オープンサイエンスと人文学・DH

3-1. オープンサイエンスとは何か

近年、「オープンサイエンス (Open Science)」という言葉が、研究政策や学術コミュニケーションの文脈で広く用いられるようになってきている。オープンサイエンスとは、研究成果を単に「無料で公開すること」を意味する概念ではない。むしろ、研究の成果が第三者にとって**検証可能であり、再利用可能である形で提示されることを重視する考え方**である。ここで言う「成果」には、論文だけでなく、研究データ、メタデータ、分析手順、用いたツールや処理スクリプト、場合によっては研究過程における判断の履歴までが含まれる。すなわち、研究を結果だけが示されるブラックボックスにせず、他者がその根拠を辿り、同じ前提条件のもとで確かめ直したり、別の目的に転用したりできる状態を整えることが、オープンサイエンスの中心的な要請である。

3-2. 人文学研究におけるオープンサイエンスの意味

この考え方は、人文学研究とも深く関わっている。人文学では、一次資料の読解や解釈が研究の中心を占め、同じ資料に対して複数の立場や解釈が並存することが前提となる。そのため、自然科学における実験再現のような意味での「再現性」を、そのまま当てはめることはできない。しかし、研究の前提となる資料の範囲、採用したテキストの形、注記や同一性判断の基準、どのような解釈上の選択が行われたのかが明示されていれば、他者はその前提を理解したうえで議論を評価し、別の解釈を提示し、あるいは同じデータを用いて異なる問いを立て直すことができる。この意味で、人文学におけるオープンサイエンスとは、単純な「再現」ではなく、**研究判断の根拠が追跡可能であり、再利用できる状態を確保すること**として理解するのが適切である。

3-3. オープンサイエンスと TEI

DH は、このようなオープンサイエンスの要請と特に親和的な研究領域である。DH の研究では、資料のデジタル化や構造化、分析のための前処理や変換、検索や可視化といった工程が不可避であり、その過程で多くの研究判断が蓄積される。成果を論文だけで提示した場合、こうした判断の多くは過程の中に埋もれてしまうが、データや処理方針を併せて提示することで、研究は他者に開かれ、検証や再利用が可能な形になる。

この文脈で TEI データの意義を捉え直すと、その位置づけは明確になる。TEI に基づいて作成

されたテキストデータには、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まとめたのか、どの典拠を参照したのかといった研究判断が、データ構造として固定化されている。すなわち、TEI データは単なる結果の保存ではなく、**研究判断を追跡可能な形で残すための器**である。したがって、TEI によるデータ作成は、研究成果を「読む」だけでなく「確かめ、使い直せる」形にするという意味で、オープンサイエンシ的な研究実践と強く結びついている。

3-4. オープンサイエンスにおける FAIR 原則

オープンサイエンスの議論と密接に関わっている概念として、FAIR 原則がある。FAIR 原則とは、研究データや研究成果が **Findable (見つけられる)** , **Accessible (アクセス可能である)** , **Interoperable (相互運用可能である)** , **Reusable (再利用可能である)** という四つの条件を満たすべきである、という考え方である。ここで重要なのは、FAIR 原則が「すべてのデータを無条件に公開せよ」という要求ではない点である。むしろ、研究成果がどのような条件のもとで利用可能なかが明確に示され、第三者がその前提を理解できることを重視している。

この観点から見ると、オープンサイエンスは、成果の全面公開を一律に求めるものではない。実際には、資料の権利関係、研究倫理、共同研究契約、個人情報などの制約により、本文テキストや画像データをそのまま公開することが難しい場合も少なくない。しかしそのような場合であっても、研究全体を閉じたものにする必要はない。たとえば、メタデータ、記述方針、用いた典拠情報へのリンク、内部リストの構造、スキーマ定義、変換スクリプトなどは、制約を受けにくく、比較的公開しやすい情報である。

3-5. FAIR 原則と TEI による記述の関係

FAIR 原則の四つの観点から見ると、TEI によるデータ記述は、研究成果を FAIR な形で提示するための実践と強く結びついている。TEI に基づくデータは、標準化された構造とメタデータを備えているため、研究成果が **見つけられる (Findable)** 状態を確保しやすい。また、公開範囲や利用条件を `<teiHeader>` 等で明示することで、**アクセス条件が明確である (Accessible)** という要件にも対応できる。

さらに、TEI は国際的に共有された語彙と構造を持つため、異なるプロジェクトやツールとの間でデータを組み合わせて利用しやすく、**相互運用可能である (Interoperable)** という条件を満たしやすい。加えて、記述方針や同一性判断の基準、参照関係が明示されていれば、他者がその前提を理解したうえでデータを新たな研究目的に用いることができ、**再利用可能である (Reusable)** 状態が実現される。

このように、TEI によるデータ作成は、オープンサイエンスと FAIR 原則の双方に照らして見た場合にも、研究成果を追跡可能かつ再利用可能な形で提示するための有効な方法である。重要なのは、どこまでを公開し、どこからが制約されるのかを明示しつつ、公開可能な部分については積極的に共有するという設計思想である。そのような姿勢によって、研究成果は個人の作業に

とどまらず、共同的に蓄積され、将来の研究へと引き継がれていく。

4. 研究発表として必要な観点

4-1. 論文として提示する場合

TEI を用いた研究成果を論文として発表する場合、最終的な成果物は従来どおり文章として提示されることが多い。しかし、その際に評価の対象となるのは、結論や解釈だけではない。どのような資料を対象とし、どのような構造判断を行い、その判断がどのように分析や議論を支えているのかが、明示的に説明されている必要がある。

TEI による構造化を行っている場合、論文では、単に「データを作成した」と述べるのではなく、どの要素をどのような基準で切り出したのか、人名や地名の同一性をどのような方針で判断したのか、内部リストと外部典拠をどのように使い分けたのかといった点を、研究方法の一部として位置づけて記述することが求められる。これにより、読者は分析結果だけでなく、その前提となる判断の枠組みを理解したうえで、議論を評価できるようになる。

この意味で、TEI による記述は論文の背後に隠れた作業ではなく、論文が依拠している方法論の一部として、適切に言語化されるべき対象である。

4-2. データとして公開する場合

一方、DH の文脈では、研究成果が論文だけで完結するとは限らない。TEI に基づいて作成されたテキストデータそのものを、研究成果として公開することも、重要な発表形態の一つである。

この場合に問われるのは、データが単に存在しているかどうかではなく、そのデータが他者にとって理解可能であり、再利用可能な形になっているかという点である。対象資料の範囲や選択理由、構造化の方針、同一性判断の基準、内部リストや外部典拠との関係などが、データや付随するメタデータを通じて説明されていなければならない。

TEI の `<teiHeader>` に記述される情報や、内部リストの設計、XInclude による参照構造は、こうした説明責任を果たすための重要な手段である。データ公開においては、どのような判断がなされ、その結果としてどのような構造を持つデータが作られたのかが追跡できることが、研究成果としての価値を支える。

4-3. 方法論として共有する場合

さらに、TEI を用いた研究では、個別の資料分析やデータ公開を超えて、方法論そのものを共有するという発表のあり方も重要となる。この場合、研究の主たる貢献は、特定の史料についての新解釈ではなく、どのような記述方針や構造設計が有効であったのか、その判断がどのような問題意識に基づいているのかという点に置かれる。

たとえば、人名や地名の同一性をどの段階で内部リストに確定させ、どの段階で外部典拠と接続するのか、あるいは、XInclude を用いたリスト管理をどのような規模や条件で採用するのが適切かといった知見は、他の研究者にとっても再利用可能な方法論的成果となる。

このような発表では、成功例だけでなく、試行錯誤や判断の揺れ、採用しなかった選択肢についても言及することが重要である。なぜその方法を選び、なぜ別の方法を採らなかったのかを明示することによって、方法論は初めて共有可能な知識となる。

このように、TEI を用いた研究成果は、論文としての発表、データとしての公開、方法論としての共有という異なる形を取りうる。重要なのは、どの形で発表するにせよ、TEI による記述に含まれる研究判断を明示し、その成果がどの水準で評価されるのかを意識的に設計することである。本書で扱ってきた演習は、こうした多層的な研究発表に耐えうる形で判断を記述し、成果として提示するための基礎を与えるものである。

5. 研究成果をどこで発表し、どのように評価されるのか

5-1. 学会・国際会議という発表の場

TEI を用いた研究成果は、まず学会や国際会議という場で発表されることが多い。デジタル・ヒューマニティーズ (DH) 分野においては、ADHO (Alliance of Digital Humanities Organizations) が主催する年次学術大会 **Digital Humanities 国際会議 (DH Conference)** が、最も代表的な発表の場である。この国際会議では、個別分野に閉じた研究成果だけでなく、方法論や研究基盤に関する発表が広く受け入れられており、TEI を用いたテキスト構造化やデータ設計に関する研究も多数報告されてきた。

また、地域や言語圏ごとにも、ADHO に加盟する DH 関連の学会・研究会が活発に活動している。欧州の European Association for Digital Humanities (EADH)、北米の Association for Computers and the Humanities (ACH) を皮切りに、カナダ、豪州圏、日本、台湾、ラテンアメリカ、南アフリカ、インドなどがあり、さらに、欧州でフランス語圏、ドイツ語圏、イタリア語圏の学会が加盟し、それぞれの地域的文脈に根ざした研究発表の場を提供している。

さらに、TEI そのものに特化した発表の場としては、TEI Consortium が主催する年次会員総会 TEI Conference and Members' Meeting が挙げられる。この会議では、TEI に関わる研究発表のみならず、TEI ガイドラインの改訂に関わる提案や、具体的なプロジェクトにおける TEI の応用事例、教育やツール開発に関する報告などが集中的に議論される。

一方、日本国内に目を向けると、DH や情報技術を用いた人文学研究は、必ずしも単一の学会に集約されてきたわけではなく、複数の学会・研究会にまたがって議論が行われてきた。たとえば、**情報処理学会人文学とコンピュータ研究会**は、文献学・歴史学・言語学・博物館学・文学研究などと情報技術との接点を扱う研究会として長い歴史を持ち、テキストの電子化や構造化、

検索・解析手法に関する研究発表の重要な受け皿となってきた。現在では年3回の研究会と年1回の査読付きシンポジウムを開催している。また、知識組織化、情報資源の記述、メタデータ、情報検索といった観点から、人文学資料を含む多様な情報対象を扱ってきた情報知識学会では、テキスト構造化や典拠管理、標準化といったテーマが継続的に議論されており、TEIを含む国際標準の理論的背景や実務的課題を検討する場として重要な役割を果たしてきた。

近年では、日本デジタル・ヒューマニティーズ学会（JADH）が、国内におけるDH研究の横断的な交流の場として年次国際研究大会を開催しており、DH全般にわたる様々な研究発表のなかで、TEI実践や教育的取り組み、国際動向との接続等も積極的に議論されている。

これらの国内学会・研究会におけるテキスト構造化に関する研究発表に共通しているのは、個別資料に関する解釈そのものだけでなく、どのような方法でデータを記述し、どのような構造判断を行ったのかが研究として問われてきた点である。特にTEIを用いた発表では、「どのタグを使ったか」という表層的な説明よりも、「なぜその構造を採用したのか」「どのような選択肢を検討し、その結果として何を選んだのか」といった方法論的説明が重要視される。

国際会議であれ国内の研究会であれ、聴衆は必ずしも対象資料の専門家ばかりではない。そのため、研究内容を特定分野の内部事情として提示するのではなく、方法論として共有可能な形で説明できるかどうか、発表の評価を大きく左右する。TEIを用いた研究は、このような分野横断的な対話の場において、共通の議論基盤を提供する点に大きな特徴がある。

5-2. 論文誌における評価

研究成果を論文として公表する場合、評価はより厳密な形で行われる。人文学系の論文誌では、従来どおりの解釈や史料批判が重視される一方で、DH関連の論文誌や特集号では、データの構造設計や方法論的貢献が明示されているかどうか重要な評価基準となる。英文ではすでに多くのジャーナルが刊行されており、代表的なものには上述のADHOによるDigital Scholarship in the Humanitiesがある。日本語論文誌としても、情報処理学会では近年、隔年でこの分野の特集号を刊行しており、JADHでも日本語論文誌を刊行している。また、データそのものを掲載するジャーナルや、それを専門としたデータジャーナルも広まってきており、データであっても査読を経た成果として公表できる場を見つけることも比較的容易になってきている。

TEIを用いた研究では、論文本文だけでなく、どのようなTEIデータを作成したのか、どの範囲まで公開しているのか、再利用可能性がどのように確保されているのかといった点も、査読の過程で問われることがある。論文とデータが分離した形で存在する場合でも、両者の関係が明確に説明されていれば、研究成果として十分に評価される。

5-3. データリポジトリと公開の意義

近年では、研究成果をデータとして公開すること自体が、独立した評価対象となりつつある。TEI/XMLデータは、リポジトリに登録し、永続的識別子を付与することで、論文と同様に引用

可能な研究成果として扱われるようになってきている。

このような形で公開されたデータは、他者による再利用や再分析を通じて、新たな研究を生み出す基盤となる。その意味で、TEI データの公開は、個々の研究成果を超えて、研究コミュニティ全体への貢献として評価される。内部リストの設計や典拠情報の扱い、XInclude による参照構造といった点が丁寧に記述されていれば、そのデータは長期的な価値を持つ。

5-4. Methodological Commons としてのデジタル・ヒューマニティーズ

DH において、TEI を用いた研究成果が特定の分野や資料を越えて評価されうる背景には、**Methodological Commons (方法論の共有地)** という考え方がある。DH は、固有の研究対象によって定義される分野ではなく、人文学研究において用いられる方法、記述の枠組み、データ構造、処理手法を、分野横断的に共有・再利用することを重視して発展してきた。そのため、研究成果の価値は、個別資料についての結論だけでなく、他の研究者が参照し、検討し、応用できる方法論的資源としての側面によっても測られる。

TEI ガイドラインは、この Methodological Commons を支える中核的な基盤の一つである。TEI による記述は、特定の研究プロジェクトの内部事情に閉じた作業ではなく、構造判断や同一性判断、参照設計といった研究上の選択を明示的に外在化し、共有可能な形で提示することを可能にする。その結果、個々の研究者が行った判断は、他者によって理解・検証され、場合によっては別の研究目的のもとで再利用されうる。

このような性質を持つ TEI の実践は、単なる技術的スキルの習得や作業効率化にとどまらない。それは、人文学研究における「どのように資料を読み、どのような前提で構造化し、どの判断を確定的なものとして提示するのか」という方法論そのものを、共有地に提供する行為であると言える。TEI を用いた研究が、学会発表や論文、データ公開といった複数の形で評価されるのは、その成果が Methodological Commons において意味を持つからである。

国際的な DH のコミュニティにおいて、TEI が方法論的基盤として評価されてきたことは、その象徴的な例である。TEI ガイドライン自体が、長年にわたる実践と議論を通じて更新され続けてきたこと、そしてその枠組み全体が国際的に顕著な貢献として評価されたことは、TEI が Methodological Commons の形成に寄与してきたことを端的に示している。

本書で扱ってきた演習は、こうした Methodological Commons に参加するための具体的な方法を示すものであった。固有表現の構造化、内部リストと外部典拠の役割分担、XInclude を用いた参照設計といった実践は、いずれも研究判断を明示的に記述し、共有可能な形で提示するための基礎である。TEI を用いたテキスト構造化は、研究成果を個人の作業にとどめず、方法論的共有地へと開いていくための重要な実践であり、その意味で DH 研究の中核に位置づけられる。

5-5. 研究成果の多層的な位置づけ

以上を踏まえると、TEI を用いた研究成果は、単一の発表形態に回収されるものではない。学

会発表、論文、データ公開、方法論的提示といった複数の層にまたがって評価される可能性を持っている。どの成果をどの場で提示するのかを意識的に設計することが重要である。

本書で扱ってきた演習は、そのための具体的な判断材料と実践的な技法を提供するものであり、読者が自らの研究成果を適切な形で発表し、評価につなげていくための基盤となることを意図している。

6. おわりに — TEI 実践から研究へ

本章では、TEI ガイドラインに基づくテキスト構造化と記述が、どのような意味で人文学研究として成立し、どのような場で評価されうるのかを検討してきた。固有表現の構造化、内部リストと外部典拠の役割分担、XInclude を用いた参照設計といった実践は、いずれも単なる技術的作業ではなく、研究判断を明示的に記録し、共有可能な形で提示するための方法である。

ここで改めて強調しておきたいのは、**TEI を用いること自体は決して自明ではない**という点である。本書の前半では、TEI を用いることの技術的・実務的な利点について繰り返し触れてきたが、本章で扱ってきた「研究成果として認められ、議論される」という文脈においては、TEI を用いることにはまた別の重要なアドバンテージがある。それは、TEI がすでに国際的に長年にわたる議論と実践の蓄積を持つ枠組みであり、その上に立って議論を進めることができるという点である。

TEI を用いた研究では、記述対象や構造設計そのものを一から正当化する必要はない。TEI ガイドラインという共通の参照枠が存在するため、議論の焦点は「なぜこのような構造を採用したのか」「既存の TEI の用法とどのような関係にあるのか」「どの点を拡張・再解釈しているのか」といった、**より高次の方法論的・解釈的問題**に置かれる。このことは、研究成果を国際的な文脈で提示し、議論するうえで極めて大きな意味を持つ。

仮にこれが、日本国内でのみ用いられている独自規格や、国際的な合意形成が十分でない枠組みであったとすれば、研究発表の多くは、その規格や形式自体を説明することに費やされてしまうだろう。その場合、研究の核心であるはずの解釈や方法論にまで議論を到達させることは容易ではない。TEI を用いることは、そうした説明コストを大幅に削減し、**すでに共有された議論の地平の上で、さらに深い検討を行うことを可能にする。**

この点において、TEI は Methodological Commons の一部として機能している。TEI による記述は、論文、学会発表、データ公開、方法論的提示といった複数の形態を通じて、国際的な研究コミュニティに向けて議論可能な形で提示される。重要なのは、TEI を用いた研究成果が、論文という単一の形式に回収される必要はないという点である。それぞれの発表形態が異なる評価軸のもとで意味を持ちうるからである。

TEI による記述を「論文を書くための準備作業」としてのみ位置づけてしまうと、こうした研究的可能性を十分に活かすことはできない。むしろ、どのような判断を行い、その判断をどのよ

うな構造として固定化したのかを意識的に言語化し、必要に応じてデータとして公開することによって、研究成果はより広い文脈で評価されるようになる。その際、TEI という国際的に共有された枠組みを用いることは、**オーディエンスを国際的な場に確保し、議論を持続的に発展させるための重要な前提条件**となる。

本書で行ってきた演習は、TEI を「使えるようになる」ことを目的とするものではなく、TEI を通じて研究判断を記述し、共有し、評価に耐える形で提示するための基礎を身につけることを目的としている。読者が本書を出発点として、自身の研究対象や関心に即した記述方針を設計し、国内外の研究コミュニティに向けて成果を発信していくことを期待したい。