

Digital Criticism Platform: エビデンスベースの解釈を支援するデジタル史料批判プラットフォーム

北本 朝展 (国立情報学研究所/総合研究大学院大学)

西村 陽子 (花園大学)

デジタル史料批判とは、研究の推論過程で積み重ねる証拠や過程をエビデンスネットワークに結合し、その分析から事実を導くための方法論である。筆者らはこれまで非文字史料を中心に史料批判の新しい方法論を開拓し、シルクロードの遺跡照合などの問題に成果を上げてきた。本論文はこの方法論をより抽象化、体系化し、エビデンスネットワークをモデルとするデジタル史料批判プラットフォーム(DCP)として、Semantic Web 技術を活用して研究基盤を構築することを目指す。さらにシルクロード遺跡に関するケーススタディでエビデンスネットワークを例示すると共に、DCP が研究のオープン化の中で果たすべき役割を考察する。

Digital Criticism Platform for Supporting Evidence-based Interpretation of Sources

Asanobu KITAMOTO National Institute of Informatics / SOKENDAI

Yoko NISHIMURA (Hanazono University)

Digital Criticism is a methodology for accumulating evidences and hypotheses in the reasoning process of research and linking them into the evidence network to derive facts through the analysis of network. We have been applying this methodology mainly to non-textual sources and obtained successful results on the identification of ruins in Silk Road areas. The purpose of the paper is to achieve better abstraction and organization of the methodology, and use the evidence network as the model of Digital Criticism Platform (DCP), which is constructed as research infrastructure using Semantic Web technology. The paper also shows examples of evidence networks in case studies about Silk Road ruins, and discusses the role of DCP when the research is requested to be more open.

1. まえがき

史料批判,あるいはより一般に source criticism と呼ばれる研究プロセスは,多種多様な史料を評価し,読み,解釈するという人文学研究の基礎をなす。歴史研究では,文字史料を研究者が読むという伝統的な場面での史料批判の方法論は確立しているが,地図や写真,画像といった非伝統的史料を「読む」場合の方法論は確立していない。そこで本論文は「デジタル史料批判」を提唱し,「エビデンスベース」の概念に基づきデジタル技術を活用した史料批判を体系化する。その核となるのが,史料や概念に関する証拠や仮説を連結したエビデンスネットワークという理論的なモデルであり,これを Semantic Web 技術を用いて実装するというのが本論文の目標である。そして「デジタル・シルクロード」プロジェクトにおいて長年取り組んできたシルクロード遺跡データベースなどをケーススタディとして,デジタル史料批判を深化させ一般化する。

本論文の構成は以下の通りである。第2章はデジタル史料批判の基本的概念とエビデンスにつ

いて述べる。第3章はデジタル史料批判プラットフォーム(digital criticism platform: DCP)のコンポーネントとその役割をまとめる。第4章はDCPの理論的なモデルであるエビデンスネットワークの構造を定義し,これをシルクロード遺跡に関する史料批判に適用したケーススタディを紹介する。最後に第5章は,オープン化時代の人文学に対する展望を交えて本論文をまとめる。

なお,本論文で用いる用語は,歴史学において使われる用語[1]を参考にしつつ,筆者らが独自に定義したものである。伝統的な定義とは意味がやや異なる場合もあるが,可能な範囲で既存の用語の意味と整合するように努力した。

2. デジタル史料批判の基本的概念

2.1 人文学における批判とエビデンス

史料を批判的に読む態度を表す言葉「source criticism」,あるいは一般的な意味での criticism や critique は,歴史学,考古学,美学,宗教学,ジャーナリズム,法律学,メディア研究に至る,人文学の広大な分野における基本的概念である。人文学における批判を考え直す提案としては,文

学におけるアルゴリズム的批判 (Algorithmic Criticism) 等の提案があり, 人間ではなく機械が「読む」ことで生じる厳密性が, 批判に新たな可能性を開くとの主張が展開されている[2].

筆者らがこれまで提案してきた「データ史料批判」はこうした方向の研究の一つと位置づけることができる. その一例にシルクロードの遺跡照合に関する研究があり, そこで鍵となったのは地図や写真等の空間画像史料の正確性に対する「疑い」であった. 空間画像史料は精細(precise)であるがゆえに正確(accurate)であるとの誤解を招きやすく, その信頼性を判定する方法も確立していなかった. そこで筆者らは, 地図や写真を「批判的に読む」方法を確立することで地図の誤差を定量的に把握する方法を確立し, 誤差を考慮すればシルクロードの遺跡が照合できることを示した[3,4].

こうした批判的な読みを支えるのがエビデンスという概念であり, 「エビデンススペース」という科学の一般原理である. 科学研究の基本は, 証拠と仮説の積み上げで形成される「エビデンスの網の目」から事実や理論を見出すという推論過程にあり, そのレベルでは自然科学と人文科学に大きな違いはない. しかし, 自然現象の客観的な観測の分析から始まる自然科学とは異なり, 人文科学では人間が作り出す文字・非文字の記録の分析から始まるという点に根本的な違いがある. そのため, 自然科学では数式一つで幅広い分野をカバーするトップダウン的なアプローチが有効な場合もあるが, 人文科学では記録を丹念に解釈してエビデンスの網の目に埋め込んでいくボトムアップ的なアプローチが不可欠となる. すなわちエビデンススペース人文科学を実現するには, ボトムアップにエビデンスを積み重ねていける研究基盤が必要である.

2.2 デジタル史料批判の提案

本論文は, 筆者らが以前から提唱してきた「データ史料批判」[5]に対して, そのアップグレード版となる概念である「デジタル史料批判」を提唱する. 以前の概念と比べると, 「データ」という史料の種類よりも「デジタル」という方法の違いに注目を移した点が異なる. そしてデジタル技術の利点を活かす観点から, ボトムアップ的な推論過程を記録するエビデンスネットワークというモデルを考案し, このモデルを核とした史料批判の体系化を進めるとの構想に至った.

ここで筆者らが「デジタル史料批判」という新しい名称を提案するのは, 史料批判に有用なデジタル技術を共通の言葉で議論する場を作りたいという意図がある. つまり, 従来は個別の技術的な方法論(アルゴリズム)として扱われてきたものを, 史料批判という歴史学の共通の俎上に載せることで, 概念を共有して議論を深化させたいのである. 歴史学は従来からエビデンスに基づく推論が人文学の中でも特に重視されてきた分野で

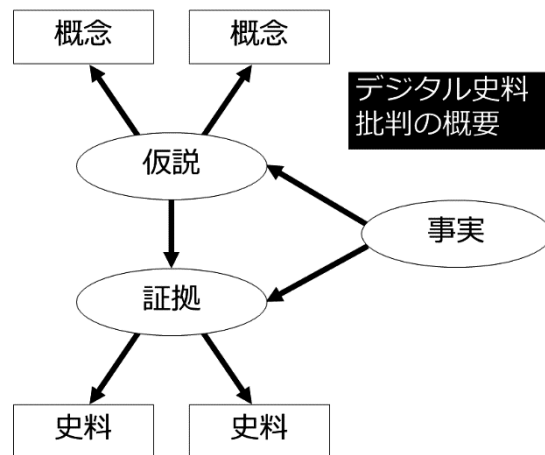


図1 エビデンスネットワークの模式図

あるからこそ, デジタル技術を用いた史料批判のモデル化によって, 研究の新たな可能性が開けることへの期待が大きいのである.

また誤解を避けるため, 人文情報学で典型的な計量的アプローチに基づく人文学と, 本論文が提唱するデジタル史料批判との違いを強調しておきたい. 計量人文学は, 史資料をデータとみなして史資料(群)から導かれる数字を分析することを目的とするが, デジタル史料批判は, 史資料を解釈した結果をエビデンスとして論理的にモデル化することを目的とする. 例えば, テキスト中の語の出現頻度や特徴的な語などを数値的に計算するのが計量人文学, 一方史資料に含まれる誤りや偏りを検証したエビデンスを論理的に結合していくのがデジタル史料批判である.

2.3 証拠・仮説・事実・確信度

エビデンススペースの史料批判を, 本論文では証拠, 仮説, 事実という3つの概念で定義し, そこに確信度という属性を導入する.

- **証拠** 史料(データ)の間の関係を指す. 例えば史料Aの一部Bと史料Cの一部Dが意味的に関連している等の知見を証拠と呼ぶ.
- **仮説** 歴史概念の間の関係を指す. 例えば遺跡概念Eと遺跡概念Fには関係がある等の知見を仮説と呼ぶ.
- **事実** 今後の研究の基礎となるような確たる関係を指す. 複数の証拠と仮説をエビデンスとして明示的に登録するが, 研究者の考え方によってその判断は異なる場合もある.
- **確信度** 人文科学では客観的な観測に基づく確かな証拠や仮説を得ることが難しく, 証拠や仮説には史資料に起因する制約が残る場合もある. そこで証拠や仮説に確信度という属性を設け, 証拠や仮説の確からしさを数字やカテゴリで表現する.

図1は、証拠、仮説、事実をノードとするエビデンスネットワークの模式図を示す。そして以下のような研究の推論過程を通して、エビデンスネットワークの構築が進むと想定する。

- (1) 様々な史料に対する証拠と様々な概念に対する仮説を結合したエビデンスネットワークを成長させる。
- (2) エビデンスネットワーク上で矛盾する証拠や仮説を見つけ出す。
- (3) 内容の再検討を通して証拠や仮説のエビデンスを洗練させ、確信度を増加させる。
- (4) 今後の研究の基礎となる事実を確定させる。さらに確信度が高い証拠や仮説だけを選択して史料批判を行うなど、動的な推論過程を実現することもデジタル環境ならではの利点である。このような推論過程を支える研究基盤とツールを、多数のケーススタディを通して洗練させることが、デジタル史料批判の重要な課題である。

3. デジタル史料批判プラットフォーム

3.1 DCPのコンポーネント

デジタル史料批判プラットフォーム DCP は、図2に示す3つのコンポーネントから構成される。

1. **データリポジトリ** デジタル史料批判の素材となる史料（データ）や、仮説の基礎となる概念、そして史料や概念に対するエビデンスを蓄積する。データはテキストに限らず画像や地図なども含み、これらに一意的識別子を与えた上でメタデータを付与し、証拠ツールや探求ツールからの利用を可能とする。
2. **証拠ツール** 史料（データ）を解釈した結果を証拠として登録するツールである。例えば2枚の写真に撮影されている遺構が同一のものであるという解釈や、2枚の地図に記載されている別名の遺跡が同一のものであるという解釈などを証拠として登録する。史料タイプの特性に合わせてツールをデザインし、解釈の結果だけでなく、状況や根拠もデータリポジトリに登録できるようにする。
3. **探求ツール** その時点で得られているエビデンスネットワークを分析し、そこから導ける仮説や事実を登録するツールである。エビデンスネットワーク上に積み上げた証拠や仮説の信頼度と、それらの整合性を考察しながら、エビデンスネットワークを洗練させていく。そのためには、エビデンスネットワークを自在に操作する機能が必要となる。

3.2 データリポジトリ

データリポジトリが提供すべき機能は以下の通りである。



図2 デジタル史料批判プラットフォーム(DCP)の模式図

- (1) 史料や概念に識別子を与え、メタデータを付与して登録する。
- (2) エビデンスネットワークを管理し、追加するエビデンスを意味的に連結可能とする。
- (3) エビデンスネットワークを意味的に検索可能とする。
- (4) 証拠ツールと連携しながら、解釈の結果とその根拠に識別子を与えて登録する。
- (5) 探求ツールと連携しながら、仮説や事実とその根拠に識別子を与えて登録する。

このような機能を実現するソフトウェア基盤として、各種のオープンソースソフトウェアを比較検討したところ、DSpaceを基盤システムとして採用することとした。その理由は、多種多様なリソースのリポジトリとして豊富な機能を有するだけでなく、Linked Data 拡張を加えることでSPARQLによる意味的検索にも対応できる点がある。また長年の開発実績は、プラットフォームの持続性の高さとして評価できる点である。

3.3 証拠ツール

証拠ツールは、対象とする史料タイプごとにツールをデザインする必要があるため、これまでに3種類の証拠ツールを開発した。

- (1) **マッピング** 地図を対象とした空間的な照合
- (2) **フォトフィット** 写真を対象とした視覚的な照合
- (3) **メモリーハンティング** 写真と実世界を対象とした空間的かつ視覚的な照合

1と2はウェブアプリ、3はモバイルアプリ[3]であり、以下では証拠ツールごとに、目的と現状をまとめる。

3.3.1 マッピング (Mapping)

マッピングは、複数の地図の照合を支援するウェブツールであり、位置ずれが大きな地図を重ね合わせて地点の対応関係を探る問題を扱う。きちんと測量された地図であればこの重ね合わせは難しくなく、測量されていない地図でも対応点（地上制御点）を拾えば、幾何補正（ジオレファレンス）によって重ね合わせは可能である。ところが古地図では、幾何補正が最適な解決策ではない場合がある。あまりに不正確な地図、あまり

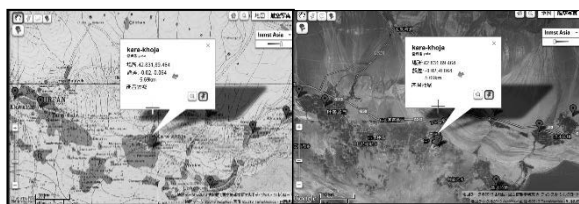


図3 マッピング: Kara-Khoja と高昌故城の同一性をエビデンスとした例.

に距離と方位が歪んでいる地図では、幾何補正に伴う地図の変形によって、肝心の地図上の情報内容が読み取れないという副作用が生じるからである。ゆえに古地図を対象とする場合は、幾何補正を使わない重ね合わせツールも必要になる。

そこで提案するのが、対話的位置合わせというアイデアに基づくマッピングである。対話的位置合わせとは、2枚の地図を利用者の指定点を基準に重ね合わせ、それ以外の部分は単なる平行移動で済ませる方法である。正確な重ね合わせは指定点に限られるが、その近傍でも誤差は十分に小さくなるため、ローカルな空間で位置関係を解釈するならこれで十分とも言える。筆者らは北京の古地図を幾何補正する手法[6]を以前に提案したが、幾何補正には地図の全体像が理解できるといったメリットと、文字が読めなくなるといったデメリットがあり、両者は使い分けが必要であるとの知見を得た。史料批判という目的では、幾何補正という全域的な位置合わせは必ずしも重要ではなく、むしろ地図上の情報内容が読める対話的な位置合わせの方が有用な場合も多いのである。

マッピングとはマップ+ピン（ピン留め）に由来し、ピン留めした地点を中心に2枚の地図を位置合わせできる機能を端的に表現した命名である。図3は古地図上の Kara-Khoja と衛星画像上の高昌故城とをピン留めした結果を示す。このようにピンを留める行為は、2地点の同一性に関する証拠を入力する行為ともみなせる。つまりデジタル史料批判のコンポーネントとして見れば、マッピングは空間史料の関連性を証拠として入力する役割を果たすものである。位置合わせパラメータとその根拠を証拠としてデータリポジトリに登録することで、誰でもその解釈の正当性を後から検証できるようになる。

3.3.2 フォトフィット (Photofit)

フォトフィットは、複数の写真の照合を支援するウェブツールであり、異なる時期に撮影された写真を重ね合わせて物体の対応関係を探索する問題を扱う。写真は景観の変化を探るための強力なメディアであるが、撮影者も時期も異なる写真が全く同一地点から同一構図で撮影されることはないため、写真がそのまま重なることはない。



図4 フォトフィット: Murtuq III Anlage と烏江不拉克烽火台の同一性を、対応点(矢印)で確認してエビデンスとした例.

つまり、ある物体を異なる地点から撮影した写真を重ね合わせるには、写真間の対応点が重なるように写真を幾何変換するツールが必要となる。

図4は、約100年前の探検隊が撮影した Murtuq III Anlage の古写真と、現代の烏江不拉克烽火台を撮影した写真をフォトフィットに読み込み、拡大縮小と平行移動を調整しながら両者を重ね合わせた結果である。2枚の写真は撮影範囲も異なり、そのままの大きさでは同一性に気づきにくい、写真を実際に重ね合わせて見ればその同一性は明瞭に把握できるようになる。位置合わせパラメータとその根拠を、証拠としてデータリポジトリに登録することで、誰でもその解釈の正当性を後から検証できるようになる。

3.3.3 メモリーハンティング (Memory Hunting)

メモリーハンティング (メモハン) [7,8]は、画像史料と実世界の照合を支援するモバイルアプリであり、画像史料が紐づく実世界上の位置と方向を探索する問題を扱う。フォトフィットは、同一地点・同一構図でない写真を撮影後に照合するツールであるが、メモハンでは写真と実世界の照合をフィールドにて撮影時に行う点に違いがある。

メモハンでは、照合すべき画像史料をスマートフォンカメラのファインダー上に半透明で重ね、実世界を歩き回りながら風景との一致度が最高となった瞬間にシャッターを押すことで、撮影地の緯度経度をスマートフォンの測位機能で記録する機能を備える。そして同一地点・同一構図の写真を撮影することで、過去と現在の景観変化も記録できるという、一石二鳥の機能を備える。

画像史料と実世界を照合するという作業は、画像史料によって難易度が異なる。第一に、写真中の著名な建造物が今も同一地点にある場合は、照合が比較的簡単である。第二に、写真のキャプション



(a) 東京での古写真照合 (対象が著名ではなく、キャプションも曖昧さを含む例)。



(b) 北京での古写真照合 (対象は著名ではないが、キャプションは明確な例)。

図 5 メモリーハンティングによるフィールドワークの例。

ョンに撮影地の明確な記述がある場合は、撮影地周辺まで到達することは容易であるため、たとえ特定が困難そうに思える場合でも、現地では手がかりが見つかって撮影地の特定が可能となる場合が多い。第三に、景観が著名でなくキャプションも曖昧な場合は、土地勘や専門知識がなければ特定に至らず、優れた特定能力を備えた人々を照合活動に巻き込むことも必要となってくる。

こうした活動は、不特定多数の参加による古写真クラウドアノテーション(crowd annotation)の試みと言える。例えば、不特定多数の人々が参加する町歩きイベントを開催し、参加者が歩き回りながら写真の撮影地を特定していくことで、フィールドワークに基づく写真アノテーションを実現する。このようにフィールドにて多人数が同時にスマートフォンで参加できる点がメモハンの特徴である。現在は Android 版アプリを無料配布しているが、将来的に iOS 版も追加してマルチプラットフォーム対応とすることで、クラウドソーシング型デジタル史料批判への展開を探る。

3.4 探求ツール

探求ツールは、証拠ツールを用いて成長させたエビデンスネットワークから導ける仮説や事実を登録するツールである。本論文で提案するエビデンスネットワークは有向グラフであり、証拠や仮説を結ぶリンクに関係性という意味を付与することから、本論文では Semantic Web 技術の一つ

である RDF (Resource Description Framework) を用いてエビデンスネットワークを構築する。RDF を用いる理由は、周辺技術が十分に成熟しており、運用レベルで利用できる品質の高いオープンソースライブラリが数多く開発されている点にある。さらに RDF のための検索言語 SPARQL を利用すれば意味的な条件を指定した検索も可能となる。これを用いて信頼度が高いエビデンスだけを抽出できれば、より確からしい証拠や仮説だけを用いて、エビデンスネットワークの分析を進めることも可能となる。

ただし一般的に設計されたツールをエビデンスネットワークという特定の目的に適用するには、独自に構築すべき理論とツールが存在する。例えば、エビデンスネットワークの検索で人文学者が SPARQL クエリを直接書く状況は想定しづらいたことから、グラフィカルユーザインタフェース (GUI) 等を活用した作業支援が必要となる。また有向グラフという構造自体が一般の人々には理解しづらいモデルであり、無秩序にグラフが成長して副作用が生じないように、研究ワークフローの工夫も必要となる。

3.5 証拠ツールの一般化

証拠ツールは、これまでに述べた 3 種類のツールに限定されるものではなく、様々な方向に一般化することが可能である。第一に、より多くの史料タイプの証拠を入力可能とする方向がある。空間史料や画像史料に加えて文字史料の照合ツールとして、例えばある史料に出てくる文字列と別の史料に出てくる文字列が同じものを指すという関係を、文字史料の一部を指定してリンク可能なツールが構想できる。第二に、異なる史料タイプをまたがる証拠を入力可能とする方向がある。例えば「高い塔がある」というテキストと高い塔の写真を対応させ、それらが同一であると解釈する証拠も扱えるようなツールが構想できる。

このような一般化の先にあるのは、任意の史料間で照合を可能とするような一群の証拠ツールである。それぞれのツールは固有の操作体系を持つが、重要なのは証拠ツールを通してエビデンスネットワークに登録する際に、エビデンスのフォーマットを共通化することである。これにより、エビデンスノード上における推論過程は外形的には統一的な扱いが可能となって、研究基盤上でのモデル化も容易となるのである。

4. エビデンスネットワーク

4.1 エビデンスネットワークの構造

図 1 に示したように、エビデンスネットワークは証拠、仮説、事実という 3 種類のノードが結ばれ

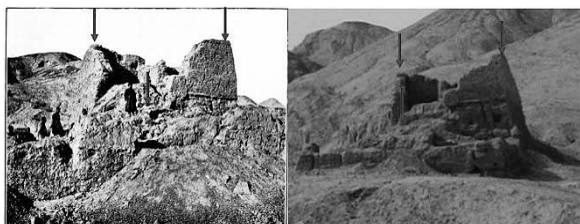


図6 Murtuk Ruins M. B. I (左) 烏江不拉克仏塔 (右) を照合した例.

た有向グラフであるが、グラフの全体を見ればさらに多くのノードが存在するため、以下の4種類のノードに分けて説明する。

1. **ソースノード** デジタル史料に対応するノードであり、識別子を用いて特定できる。
2. **インスタンスノード** 歴史を記述する単位に対応するノードであり、識別子を用いて特定することができる。インスタンスとは、抽象的な概念(クラス)の実例(インスタンス)という意味である。例えば「遺跡」という概念に対する個々の遺跡がインスタンスノードに相当する。類似する用語に実体としてのエンティティという言葉があり、例えば歴史上の有名人などはエンティティと呼んでも差し支えない。ただし一般の場合は、何を単位として論じるかという概念化[9]そのものが研究の対象となり、実体が先に存在するわけではないという意味で、インスタンスという用語を用いることとした。
3. **エビデンスノード** 証拠および仮説に対応するノードであり、それぞれ確信度という属性を用いて確からしさを表現する。これらは史料の批判的な検討という研究の推論過程の中核をなすものであり、エビデンスノードとその間のリンクが研究の成果となる。
4. **事実ノード** エビデンスネットワークから、確実に成立する関係性や、研究者のコンセンサスが得られる関係性を見出し、それらを事実ノードとして登録する。今後の研究において安心して引用できるレベルの関係性を対象とすることから、登録にはエビデンスノードよりも慎重な検討が必要である。

ソースノード以外、特にエビデンスノードにおいては、何を証拠として認定するか、何を仮説として提示するかには、個人差があることが予想される。そこで各ノードに入力者IDを付与し、誰がその証拠や仮説を登録したかという情報を保持すると共に、入力者が自らの判断に基づきエビデンスの確信度を入力できるようにする。確信度は証拠や仮説の信頼性の判断材料となるだけでなく、SPARQLによる検索で確信度が高いエビデンスのみを抽出するなど、エビデンスネットワー

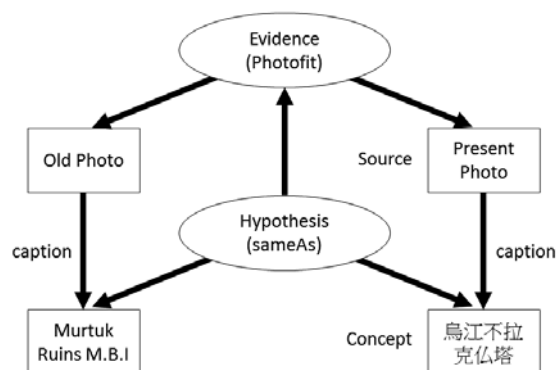


図7 Murtuk Ruins M. B. Iと烏江不拉克仏塔に関連するエビデンスネットワーク。

クを洗練化させる場面でも利用する。なお、確信度は事実ノードにも入力可能とするが、事実ノードは本来なら確実なものや研究者のコンセンサスが得られるものに限定すべきであるため、確信度は積極的に利用しないこととする。

4.2 エビデンスネットワークの例

以下では、筆者らがすでに出版した論文[3]の事例を対象として、エビデンスネットワークの構築方法を例示する。これらのエビデンスネットワークは、実際に研究が進行した過程に対応するというよりは、様々な試行錯誤を経て得られた研究成果を、エビデンスネットワークというツールを使って再現したものである。また本論文では史料批判の中身については議論せず、史料批判の結果として得られるエビデンスネットワークに焦点を合わせる。史料批判の具体的な中身については、既出版の論文[3,4,5]に記述を譲ることとする。

4.2.1 Murtuk Ruins と烏江不拉克仏塔

図6はMurtuk Ruinsと烏江不拉克仏塔を撮影した写真を示す。ここでの出発点は、ある古写真中の「Murtuk Ruins」という名前の遺跡と、ある現代写真中の「烏江不拉克仏塔」という遺跡が同じものではないかという推定である。写真と遺跡名は写真のキャプションを通して結ばれているが、遺跡名には綴りの類似性がないため、名寄せだけで同一性を判定することは不可能である。しかし、古地図を使った場寄せを進める過程で、これらの遺跡の同一性に関する探究を進めることになった。そこでフォトフィットを使って2枚の写真を比較すると、写真を平行移動・拡大縮小することで、写真中の遺跡形状の対応付けが可能であることが判明した。このように写真に対する史料批判を通して、2つの遺跡の同一性に関する仮説を提示することができた。

以上の推論過程をエビデンスネットワークと

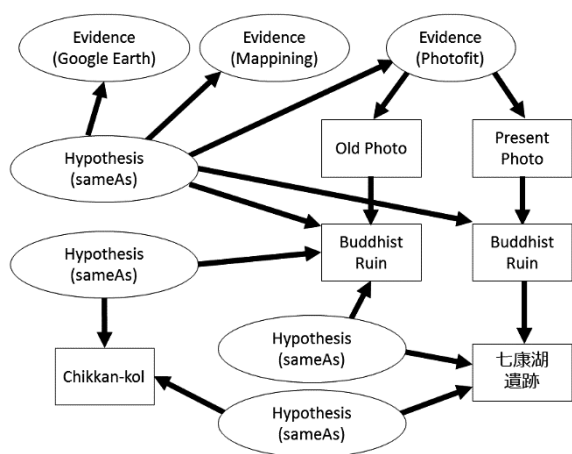


図 8 Chikkan-köl と七康湖遺跡に関連するエビデンスネットワーク。

して表現したのが図 7 である。四角ノードはソースノード、楕円ノードはエビデンスノードであり、それらを有向グラフで結んだものがエビデンスネットワークとなる。まずは写真のキャプションを用いて、史料と遺跡をリンクする。次にフォトフィットを用いて写真の重ね合わせを行い、史料上での一致という証拠をエビデンスノードとして登録する。そして、フォトフィットで得られた証拠を基に、2つの遺跡が同一のもの (sameAs) であるという仮説をエビデンスノードとして登録する。これにより、この研究者が 2 枚の写真の照合という解釈から導いたエビデンスネットワークが完成した。今後の現地調査などでさらにエビデンスを蓄積できれば、この 2 つの遺跡の同一性を事実と認定し、今後の研究に引用していくことが可能になるかもしれない。

4.2.2 Chikkan-köl と七康湖遺跡

次に、より複雑なエビデンスネットワークとして、複数の史料を照合して証拠を積み上げ、それらの証拠を基にして遺跡の同一性に関する仮説を立てた場合を述べる。出発点は、ある仏教寺院遺構と別の仏教寺院遺構とが同一のものではないかという直観である。そこで、マッピングを用いた地図上での「場寄せ」の証拠、フォトフィットを用いた写真照合の証拠、そして Google Earth を用いた景観照合の証拠などを次々と登録することで、エビデンスネットワークを成長させていく。一方これとは別に、仏教寺院遺跡と七康湖遺跡との同一性に関する証拠が得られており、さらにこの仏教寺院遺跡と Chikkan-köl として知られている場所の同一性に関する証拠が得られている。そこでこれらがお互いに同一であるという仮説を立ててエビデンスネットワークに登録して図 8 のネットワークを得た。これにより、3つ

```

PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX dcp: <http://dsr.nii.ac.jp/dcp#>
SELECT ?instance_a ?Relation ?instance_b ?CF
WHERE {
  ?Hypothesis
    dcp:certaintyFactor ?CF ;
    dcp:subjectUri ?a ;
    dcp:objectUri ?b ;
    dcp:relation ?Relation .
  ?a dcterms:title ?instance_a .
  ?b dcterms:title ?instance_b .
  FILTER (?CF > 80)
}
    
```

(a) SPARQL を用いたクエリ。確信度の高いエビデンス (CF>80) のみを検索した。

instance_a	Relation	instance_b	CF
"仏教寺院遺跡 (Hotcho)"	<http://schema.org/sameAs>	"Chikkan-köl"	85
"仏教寺院遺構 (Hotcho)"	<http://schema.org/sameAs>	"仏教寺院遺構 (現代地理調査)"	90
"Chikkan-köl"	<http://schema.org/sameAs>	"七康湖遺跡"	90

(b) SPARQL 検索結果。今後はこれをグラフとして可視化する計画である。

図 9 エビデンスネットワークの SPARQL 検索例。

の遺跡と地点が同一であるという仮説が得られ、エビデンスネットワークを再検討しても個別の証拠や仮説に問題はなさそうであると判断できたことから、七康湖遺跡と Chikkan-köl は同一の地点であるという、遺跡や地点に関する未知の関係性を仮説として提示することができた。

最後に、このエビデンスネットワークを対象とした SPARQL 検索の例を図 9 に示す。これは確信度が高いエビデンスのみを検索するというクエリであるが、これにより検討対象とするエビデンスを動的に選択できるため、エビデンスネットワークの分析や洗練化をより効率的に行えるようになると思われる。

なお図 7 や図 8 で示したエビデンスネットワークは、研究の再現として構築した仮想的なものであり、実際の研究をこのツールで推進した例はまだ存在しない。今後は多数のケーススタディにおいて DCP を実利用し、研究者のフィードバックに基づき DCP を改良する計画である。

5. あとがき

本論文は、デジタル技術を活用した史料批判の方法としてエビデンスネットワークというモデルを提案し、それを実装したデジタル史料批判プラットフォームの現況について報告した。証拠と仮説を積み上げて事実を導くという方法論自体は、既存の歴史学の研究方法与大きな違いはない。逆に言えば、既存の歴史学の方法と違和感の

ない形でワークフローを定義すれば、デジタル化された史料批判のメリットを最大限に引き出すことができる。デジタル技術を梃子にして歴史学の可能性を広げるために、研究者にとって違和感のない研究ワークフローが利用できる研究基盤を構築することが DCP の課題である。

本論文は歴史学を対象を絞って議論してきたが、DCP と同様のアイデアは人文学の他の分野でも活用できると考えている。人文学における批判をデジタル史料批判という概念で理論化すると共に、その実装としてのシステムを各分野に広げていく方法が望ましいと考えている。

最後に、このプラットフォームの長期的な価値として、エビデンスサイテーションの構想について述べておきたい。様々な学術分野において、研究のエビデンスとなるデータの公開が求められるようになってきた。その背景には、研究成果の再利用を促進するという利便性の側面と、研究成果の不正を防止するという透明性の側面とがあり、より広い視点で見れば研究のオープン化を求める社会の声を反映した動きと捉えることができる。オープン化の程度は分野によって異なるとはいえ、人文系にもこうしたオープンサイエンス [10] の動向が波及してくることは確実である。このとき人文系において、エビデンスとなるデータは何かという問題を考える。

まず研究資源のオープン化が課題となる。限られた研究者しかアクセスできない秘匿史料を分析した結果は信頼できないという評価が定着すれば、史料のデジタル化とオープン化は必須の要件となろう。次に研究成果のオープン化が課題となる。研究プロセスの透明化のために、成果の根拠となるデータも含めてオープンアクセスとする手段が必要となってくる。そこで重要となるのが、(1) 研究の推論過程で利用されたエビデンスを、誰もが検証可能な形で公開することができ、(2) エビデンスを他者が再利用して新たな知見を付け加えることで、エビデンスネットワークが成長する可能性も秘めた、デジタル史料批判プラットフォームの役割である。つまりデジタル史料批判プラットフォームとは、オープン化に向かう研究の動向を先取りした研究基盤であるとの捉え方も可能である。

このような世界では、論文という形式から漏れていた小さなエビデンスもデジタル出版が可能となり、出版されたエビデンスに対する引用(サイテーション)を通してエビデンスの価値が高まることで、細粒度の研究成果を流通させる道も開けてくる。これは、推論過程を網羅的に記録するエビデンスネットワークのような仕組みを用いることで初めて実現できる世界である。

もちろんすべてをオープン化すればよいわけではない。価値の高いエビデンスは論文が採録されるまで非公開にするなど、研究成果の保護に関する機能は必要である。とはいえ、秘密の史料と

曖昧なエビデンスに基づく閉じた人文学から、オープンな史料と明示的なエビデンスに基づく開かれた人文学へというトレンドは変わりそうもない。そうした変化を支援する研究基盤という役割が、デジタル史料批判プラットフォームの長期的な価値につながると考えている。

謝辞

デジタル史料批判プラットフォームの構築には、ソフトウェアエンジニアの池崎友博氏の協力を得た。本研究は、JSPS 科研費 26540178 「データ史料批判：非文字史料の情報学的解析に基づくシルクロード像の再構築」の助成を受けた。またメモハンの構築の一部には、京都大学地域研究統合情報センター共同研究の助成を受けた。

参考文献

- 1) 遅塚 忠躬, “史学概論,” 東京大学出版会, 2010.
- 2) Stephen Ramsay, “Reading machines: toward an algorithmic criticism,” University of Illinois Press, 2011.
- 3) 西村 陽子, 北本 朝展, “スタイン地図と Google Earth を用いた名寄せと場寄せに基づくシルクロード探検隊調査遺跡の解明,” 人文科学とコンピュータシンポジウム じんもんこん 2010, pp. 255-262, 2010.
- 4) 西村 陽子, 北本 朝展, “地図史料批判に基づくシルクロード都市遺跡・高昌故城の遺構同定,” 人文科学とコンピュータシンポジウム じんもんこん 2014, pp. 43-50, 2014.
- 5) Asanobu KITAMOTO, Yoko NISHIMURA, “Data Criticism: General Framework for the Quantitative Interpretation of Non-Textual Sources”, Digital Humanities, 2014.
- 6) 西村 陽子, 北本 朝展, “Google Earth と『乾隆京城全図』を用いた北京歴史空間の情報基盤”, 人文科学とコンピュータシンポジウム じんもんこん 2008, pp. 81-88, 2008.
- 7) 北本 朝展, “デジタル人文学：コンテンツの「解釈」を重視したメディア技術の展開”, 精密工学会 画像応用技術専門委員会 第5回定例研究会「クリエイティブ・コンテンツ：ファッション, 伝統工芸, 観光・メディア」, pp. 1-10, 2015.
- 8) Asanobu KITAMOTO, “MemoryHunt: A Mobile App with an Active Viewfinder for Crowdsourced Annotation through the Re-experience of the Photographer”, Fifth Annual Conference of the Japanese Association for Digital Humanities (JADH2015), 2015.
- 9) 溝口 理一郎, “オントロジー工学,” オーム社, 2005.
- 10) マイケル・ニールセン, “オープンサイエンス革命,” 紀伊国屋書店, 2013.