

史料デジタル収集の体系化に基づく

歴史オントロジー構築の研究

A Study of Constructing of the Historical
Ontology with Digitizing of Historical Materials

林 譲 (HAYASHI YUZURU)

東京大学・史料編纂所・教授



研究の概要

東京大学史料編纂所が60年間にわたって収集・蓄積する探訪史料マイクロフィルムをデジタル化し、ポーンデジタルによる収集の仕様を確立し、一点目録情報などのメタデータを付与したデジタルデータを格納するアーカイヴハブ（デジタル画像史料収蔵庫）を構築する。あわせて、デジタル画像史料群に基づく先端的项目・歴史オントロジー構築の研究を推進する。

研究分野：人文学

科研費の分科・細目：史学・日本史

キーワード：日本史・史料・デジタル・アーカイヴハブ・歴史知識学・オントロジー

1. 研究開始当初の背景

研究代表者らが所属する東京大学史料編纂所では、国内外の日本史史料を組織的・系統的に調査・収集し、その成果は複成本（影写本・謄写本など）として、利用されてきた。なかでも、マイクロフィルム・カメラ撮影による複成は「写真帳」として広く利用されてきたが、2008年以降、複写用印画紙、史料撮影に適した35ミリロールフィルムなどの生産が中止になるなど、マイクロフィルムとカメラによる史料探訪・複製作成システムの継続が全く困難な状況に陥った。

一方、史料編纂所では、所蔵する影写本・謄写本・史料稿本類のデジタル化（総計530万件）に取り組み、研究資源の高度情報化を進めていた。

2. 研究の目的

本研究は、これらの成果・資源を継承し、アナログ媒体のデジタル仕様による情報資源化を目指し、史料収集システムのデジタル仕様への再構築を図るものである。

その目的とするところは、第一に、探訪史料フィルム類のデジタル化とアーカイヴハブ（デジタル画像史料収蔵庫）の構築、およびデジタル撮影（ポーンデジタル）による史料収集の仕様の確立、第二に、デジタル画像史料群に基づく先端的项目・歴史オントロジー構築の研究、にある。

3. 研究の方法

まず、史料編纂所が60年間にわたって系統的に蓄積するフィルムベースの史料画像をデジタル化する。総計およそ490万コマのうち、フィルムの劣化や原本史料の欠如・公開要請の高いものなど、緊急性や必要性の高いものを当面の対象とする。あわせて、デジタル撮影（ポーンデジタル）による史料収集の仕様を研究し、モデル的な史料収集を推進する。

デジタル化した史料画像はファイルサーバに格納し、史料の原秩序と階層構造を反映させたメタデータを付与する。必要に応じて一点ごとの目録情報も作成し、サーバ空間内に巨大なアーカイヴハブ（デジタル画像史料収蔵庫）を構築し、研究資源の高度情報化と共同利用をはかる。

そして、アーカイヴハブのデジタル画像史料群をベースに、画像研究とテキスト研究における先端的项目を推進し、史料編纂所の歴史情報処理システム（SHIPS）データベースとアーカイヴハブとの統合を目指す。また、歴史知識を蓄積する歴史オントロジーの構築について研究する。

4. これまでの成果

本研究を時間軸に沿って整理すると、2008年度は課題の遂行に最適なサーバなどの基盤を整備してデータ蓄積を開始し、2009年度は画像ビューワ・検索システム構築による閲覧・検索可能な環

境を創り出し、2010年度はWebコンテンツ保護機能・HI-Cat Plus など公開とセキュリティ対策のシステム開発を行ってきた、と語り得る。

以下、具体的な成果を報告する。

- ①利用や公開を視野に入れた最適な方法を研究し、採訪史料マイクロフィルム200万コマ、カラーシートフィルム4350枚のデジタル化を完了した。
- ②史料収集時の諸情報（所蔵者、撮影年月日、史料群名など）、一点目録情報で構成されるメタデータとともに、デジタル化した史料画像やデジタル撮影（ボーンデジタル）データを格納するアーカイヴハブ（デジタル画像史料収蔵庫）を構築した。
- ③画像検索システムを構築し、キーワード検索・撮影年の範囲指定検索により、高精細画像閲覧モジュールを組み込んだ画像閲覧ビューワが立ち上がり、デジタル画像を閲覧できるようにした。
- ④画像閲覧システムを構築し、紙媒体の「写真帳」では望みえなかった画像の拡大縮小表示、表示位置の変更、前後や指定した画像への遷移、サムネイル表示、また折紙に対応する画像回転表示、紙背文字を読み易くする上下・左右反転表示など、一般的ビューワでは実現しにくい機能を実装した。
- ⑤厳密な制御が不可欠な閲覧・印刷に関して、Webコンテンツ保護システムの研究開発により、閲覧・コピー&ペースト・印刷・画面キャプチャーの諸段階にわたるグループ毎の制限を可能とした。
- ⑥デジタル撮影（ボーンデジタル）による史料収集の仕様を策定した。採訪コードを導入し、採訪から利用までの一貫した管理を可能とした。
- ⑦メタデータ検索と画像閲覧を実現するため、一史料画像毎に差出（発信者）・宛所（受信者）・発信年代・内容等を研究して史料名を付与したデジタル史料画像約40万コマの一点目録を作成した。
- ⑧一点目録検索システムとして、史料編纂所所蔵史料目録データベース（HI-CAT）と連携したHI-CAT PLUSを開発し、アーカイヴハブ格納のデジタル史料画像群の検索・閲覧が可能になった。
- ⑨HI-CATとの横断検索を実現し、目録系・フルテキスト系データベースについても、厳密なセキュリティコントロールをもつHI-CAT PLUS経由によるデジタル史料画像群との連携を可能とした。
- ⑩歴史オントロジー構築の研究に関して、アーカイヴハブ/HI-CAT PLUSシステムに、歴史＝史料知識オントロジー機能を搭載するため、採訪史料のマイクロフィルム単位のまとまりを表わす採訪プロパティの検討・策定、採訪プロパティによる実在する史料のまとまりの再現の可能性、歴史知識の概念オントロジーを採訪プロパティの相互間の

関係定義による方針の確立、について準備した。

以下は、種々と連携して推進している個別の研究プロジェクトの例示である。

- ⑪秋田藩家蔵文書デジタル収集とバーチャル「秋田藩採集文書」の構築、⑫『信長文書』フルテキストデータ構築と史料画像リンク、⑬⑭萩原寺地蔵院聖教プロジェクト・金光図書館具注暦プロジェクト、⑮電子くずし字字典データベースを軸とした機関間連携の推進、および字形・字体データから人物オントロジーへの展開に関する研究、⑯古写真プロジェクト、⑰中国档案プロジェクト

5. 今後の計画

- ①採訪史料マイクロフィルムのデジタル化を継続して行う。
- ②仕様策定に基づくデジタル撮影（ボーンデジタル）による史料収集を推進する。
- ③一点目録情報の作成を継続して行い、アーカイヴハブ（デジタル画像史料収蔵庫）を充実する。
- ④画像検索・閲覧・コンテンツ保護各システムの整備に基づく史料編纂所図書室での公開を進める。
- ⑤引き続きデジタル画像史料群に基づく画像・テキスト研究プロジェクトを推進する。
- ⑥採訪プロパティによる歴史オントロジーの構築研究を推進する。

6. これまでの発表論文等（受賞等も含む）

- ①山田太造・横山伊徳・綱川歩美・高橋典幸・林讓「採訪史料管理システム」（人文科学とコンピュータシンポジウム『じんもんこん2010』）、145-150頁、2010年
- ②横山伊徳「史料編纂とデジタル化のメタヒストリー」（『人工知能学会誌』25-1号）、5-10頁、2010年
- ③横山伊徳「日本史研究データベースはどこへ行こうとしているのか」（『日本歴史』740号）、2-8頁、2010年
- ④金子拓「秋田藩家蔵文書のデータベース化と地域連携—「秋田藩採集文書論」のために—」（『東京大学史料編纂所シンポジウム 研究と情報の資源化』予稿集）、5-25頁、2010年
- ⑤林讓「科学研究費基盤研究(S)「史料デジタル収集の体系化に基づく歴史オントロジー構築の研究」の現状と課題について」（同上）、1-4頁、
- ⑦石川徹也・横山伊徳編著、赤石美奈、遠藤基郎、近藤成一著『歴史知識学ことはじめ』勉誠出版、202頁、2009年
- ⑧村井祐樹・末柄豊編著『真如寺所蔵能勢家文書』（『東京大学史料編纂所研究成果報告 2010-1』）、70頁、2010年