

Hi-CAT Plus: デジタル史料の検索・閲覧システム

大内英範[†] 山田太造[‡] 高橋典幸[†] 綱川歩美[†]

林譲[†] 保谷徹[†] 山家浩樹[†] 横山伊徳[†]

[†] 東京大学史料編纂所 [‡] 人間文化研究機構

Hi-CAT Plus は、採訪マイクロフィルムをデジタル化した画像の検索・閲覧システムとして開発され、史料編纂所閲覧室の端末でサービスをはじめた。本システムの仕組みや既存システムとの連携、上記用途にとどまらない今後の展望などについて述べる。

Hi-CAT Plus: A searching and viewing system for digitized historical resources

Hidenori Ouchi[†] Taizo Yamada[‡] Noriyuki Takahashi[†] Ayumi Tsunakawa[†]
Yuzuru Hayashi[†] Toru Hoya[†] Koki Yanbe[†] Yoshinori Yokoyama[†]

[†] Historiographical Institute
The University of Tokyo

[‡] National Institutes
for the Humanities

Hi-CAT Plus is a searching and viewing system for digitized historical resources. The Historiographical Institute started to provide its service in its reading room at July 1st 2011. In this paper, we present a structure, cooperation with the existing system and future prospects of the Hi-CAT Plus.

1. はじめに

本年 7 月 1 日、東京大学史料編纂所（以下「史料編纂所」）閲覧室で、デジタル化された採訪マイクロフィルム（後述）の検索・閲覧サービス「Hi-CAT Plus」の運用を開始した。利用は史料編纂所閲覧室備え付けの端末に限られ、まだインターネット経由での利用等はできない。本論文では、本「Hi-CAT Plus」開発に至った経緯、仕組みと利用方法、課題と今後の予定について述べる。

2. 本研究の背景

史料編纂所は、国内外の日本史史料を収集・研究し、その成果を踏まえて史料集の編纂・出版をおこなう研究所である。そのために 120 年以上にわたり、組織的・系統的な史料の調査・収集活動をおこなってきた。この「史料採訪」（各地での史料調査・収集）は当初は史料を借り受けて書き写す方法（注：影写、謄写など。史料編纂所ではいわゆる敷き写しのことを影写といい、内容をただ書写することを謄写といっている。）がとられたが、1950 年ごろからマイクロフィルムによる撮影を中心とするようになり、現在までに約 500 万コマの採訪マイクロフィルムを撮影・蓄積してきた。さらに、撮影された史料の紙焼きによって写真帳を作成し、書庫に配架することで、当該史料を（外部の閲覧

者も含めて）共有してきた。しかし近年、紙焼きに用いる印画紙の生産中止、マイクロフィルム自体の生産も縮小してゆく中で、史料編纂所が長年にわたって築いてきた史料共有の仕組みを見直す必要に迫られた。

一方、前述の影写・謄写によって作成され、蓄積してきた影写本や謄写本などについては、それらのデジタル化による研究資源の高度情報化が既に進められており、閲覧室においては『史料編纂所所蔵史料目録データベース（Hi-CAT）』の検索結果からその画像を閲覧することができるようになっている。これらの成果・資源を継承しながら、史料収集・共有システムのデジタル仕様の確立および高度化をめざすこととなった。具体的には以下の 2 点について計画・進行中である。

(1) 採訪史料フィルム類のデジタル化とアーカイブ（デジタル画像史料収蔵庫）の構築、およびデジタル撮影（ボーンデジタル）による史料収集の仕様の確立

(2) デジタル画像史料群に基づく先端的プロジェクト・歴史オントロジー構築の研究

本論文で述べる Hi-CAT Plus は、デジタル化された史料の共有方法として、上記(1)を発展させながら同時に(2)の基盤となる、研究全体の中でも重要な位置を占めるものとして開発された。

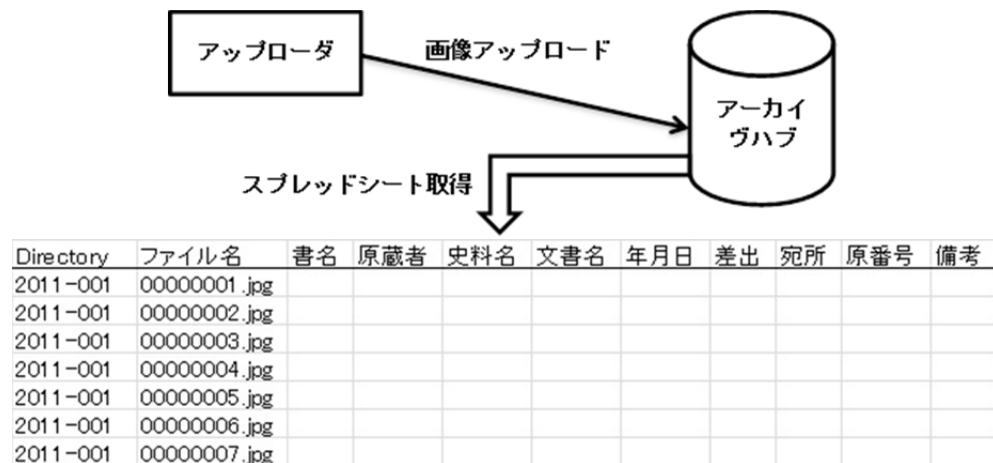


図 1. アップロードした画像ファイルについてのスプレッドシート取得. ※「Directory」および「ファイル名」にはアーカイブハブ内でのリンク情報を持たせてある.

3. 採訪史料の管理

デジタル化したマイクロフィルムの画像を管理する方法については、簡潔に述べる。

デジタル化した画像ファイルは、マイクロフィルムの 1 リールを管理の最小単位（1 ディレクトリ）とし、所蔵者・所蔵機関、撮影年、史料群名などの史料収集時の諸情報から構成されるメタデータとともに、専用アップローダによって収集資料レポジトリ「アーカイブハブ」に格納される。上記諸情報のキーワード指定による全文検索や、撮影年の範囲指定による時間検索が可能な簡易検索システムを構築し、現在 120 万コマ分の検索・閲覧が可能である。なお、現段階ではこのシステムは所員向けのものであり、所外に公開されていない。（将来的には公開の予定である。）

4. 一点目録の作成

4.1. 概要

前節で述べた管理によれば、史料画像はマイクロフィルム 1 リールを単位として、撮影年や所蔵者、史料群名などの情報とともに目録化され、簡易検索システムで検索・閲覧できる。しかし、その史料群のなかに具体的にどのような史料が含まれているかは撮影者でなければわからず、具体的な史料一点一点の利用はこれまでできなかった。そこで、前記アーカイブハブが生成する格納位置情報を示すスプレッドシートをもとに、デジタル史料画像群の一点目録を作成している。古代から明治維新に至る多種多様な史料画像が収集・蓄積されるため、その作成に際しては、研究者により一史料画像ごとに差出（発信者）・宛所（受信者）・発信年代等を分析・確定し、さらに内容を調査して史料名、文書名を付与している（史料メタデータの作成）。2010 年度までに、史料画像約 40 万コマの

一点目録作成を完了、さらに今年度 40 万コマ分の作成を終え、登録作業を準備中である。

この一点目録（を構造化したもの）が、Hi-CAT Plus の検索対象となるメタデータとなるわけである。以下、本節および次節にて、この一点目録の作成およびその構造化について述べる。

4.2. スプレッドシートの取得

アップローダにより格納された画像ファイルについて、その格納位置情報を示すスプレッドシートを生成するようにした（図 1）。格納位置情報とは、「UpperDir」（注：アーカイブハブ内のディスクマウント名）、「Directory」（注：マイクロフィルム 1 リール毎に作成されている Directory 名）、「ファイル名」である。

「Directory」および「ファイル名」に格納される値には、そのディレクトリおよびファイルにリンクが設定され、作業者によるファイルや画像の確認作業を容易におこなうことができるようとした。また、このスプレッドシートにはあらかじめ上記格納位置情報以外の項目（書名、原蔵者、史料名、文書名、年月日、差出、宛所、原番号、備考）が作成されている。

4.3. 「書名」「原蔵者」の入力

上記スプレッドシートは、1 画像ファイルを 1 レコードとして管理するものである。まずすべてのレコードに、その「書名」および「原蔵者」フィールドを入力する。これはマイクロフィルムで撮影する際に最初のコマに撮影する「ターゲット」を参照して入力する（図 2）。また、3 節に述べた簡易検索システムで検索できるのはここまでデータである。



図 2. 「ターゲット」の情報から「書名」「原蔵者」を入力

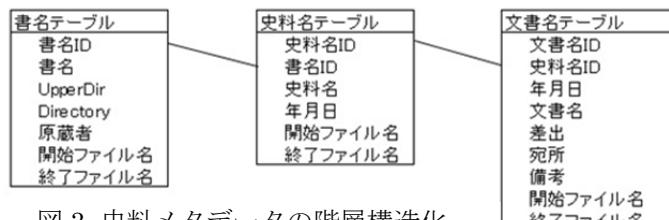


図 3. 史料メタデータの階層構造化

4.4. 「史料名」以下の入力

「史料名」以下の入力は、画像 1 ファイルずつを見ながらの作業となる。複数の文書が 1 冊にまとめられていたり、継がれて 1 卷になっていたりして、題箋などが付されている場合にはそれを「史料名」とし、さらにそれを構成している 1 つ 1 つの文書に適宜「文書名」を付していく。何らかのまとまりを構成していない 1 点ずつの文書の場合には、「史料名」「文書名」が同じになることもある。加えて、ある人物からある人物への書状が複数あり、年月日がそれぞれ特定できない場合、「差出」「宛所」まで同じであるが実態は別文書というケースもある。そのような場合はレコードの末尾に便宜的に a や b などの記号を付して区別できるようにしている。

「史料名」「文書名」「年月日」については、ともに明記されていないことも多く、作業者の判断（すなわち調査・研究）に負うところが大きい。また、「差出」「宛所」なども、1 点 1 点の文書の読み解が必要である。前記のとおり、研究者による作業によらなければ内容を確定することができない。

5. Hi-CAT Plus

前節で述べたような作業を 1 画像ファイルずつに施して完成したスプレッドシートのデータは、図 3 のように正規化（構造化）されて Hi-CAT Plus の検索用データとなる。（史料メタデータの階層構造化）

各 ID はテーブルをつなぐ外部キーとなる。各テーブルにはそれぞれの階層に応じた「開始ファイル名」および「終了ファイル名」が格納され、各「書名」「史料名」「文書名」に応じたファイルの範囲を示す。

検索はキーワードで各項目におこない、検索結果をまず「書名」レベルで一覧に表示する

（図 4）。検索結果一覧から「全体表示」を押下すれば該当「書名」全体の詳細表示、「限定表示」を押下すれば該当「史料名」全体についての詳細表示を得ることができる（図 5）。さらに詳細画面の「イメージ」ボタン押下により、高精細画像ビューアが立ち上がり、該当画像ファイルを閲覧できる。このビューアはマイクロフィルムリーダに似たインターフェースで、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、研究上必要と思われるいくつかの機能を装備したものである。また、「イメージ」ボタンは各階層ごとに配置されており、これは前述の各テーブルに格納した「開始ファイル名」「終了ファイル名」の 2 つの項目に対応している。「イメージ」ボタン押下により、その範囲がビューアに渡され、サムネイル表示におけるハイライト、「先頭」「最終」ボタンに持たせる値として用いられる（図 6）。

このように、Hi-CAT Plus では、採訪マイクロフィルムのデジタル画像を 1 点ごとに検索・閲覧できる。ここではその検索結果の一覧表示がマイクロフィルムの物理単位ではなく、「書名」単位であることを強調しておきたい。「書名」には多くの場合「○○家文書」など、（多くは 1 つの「原蔵者」の中での）史料のまとまりを示す文字列が格納されている。これは Hi-CAT Plus が、物理単位すなわち採訪活動単位に即した管理構造を持つアーカイブハブと異なり、一点一点の史料に密着した管理構造を持っていることを示している。

なお、「Hi-CAT Plus」で検索・閲覧できるこれら採訪画像は、基本的に史料編纂所以外の個人・機関が所蔵する史料の画像であり、写真帳作成およびデジタル化についての許諾を得るようしている。それらは原則として閲覧室での公開を前提としているため、現在のところ史料

検索結果(1件) 検索式: キーワード(長崎)					
ソート					
1 2 3 4 次へ 最終へ					
No	架	番	書名	原蔵者	表示種別
1	新訂マ	1992-001	長崎県立長崎図書館所蔵文書	長崎県立長崎図書館	全表示 [隠定表示]
2	新訂マ	1992-002	長崎県立長崎図書館所蔵文書	長崎県立長崎図書館	全表示 [隠定表示]
3	新訂マ	1992-003	長崎県立長崎図書館所蔵文書	長崎県立長崎図書館	全表示 [隠定表示]
4	新訂マ	1992-004	長崎県立長崎図書館所蔵文書	長崎県立長崎図書館	全表示 [隠定表示]
5	新訂マ	1992-005	寺崎文書	坂口多喜江	全表示 [隠定表示]
6	新訂マ	1992-005	武田文書	武田家幸	全表示 [隠定表示]
7	新訂マ	1992-006	余家文庫史料	長崎県立村黒歴史民俗資料館	全表示 [隠定表示]
8	新訂マ	1992-007	余家文庫史料	長崎県立村黒歴史民俗資料館	全表示 [隠定表示]
9	新訂マ	1992-008	余家文庫史料	長崎県立村黒歴史民俗資料館	全表示 [隠定表示]
10	新訂マ	1992-113	新鹿家文書	明治大学附属博物館	全表示 [隠定表示]
11	新訂マ	1992-115	新鹿家文書	明治大学附属博物館	全表示 [隠定表示]
12	新訂マ	1992-143	井伊家史料	井伊直慶	全表示 [隠定表示]
13	新訂マ	1992-144	井伊家史料	井伊直慶	全表示 [隠定表示]
14	新訂マ	1993-001	長崎文書	東洋大学文学部国史学研究室	全表示 [隠定表示]
15	新訂マ	1993-009	毛利家文庫史料	山口県文書館	全表示 [隠定表示]
16	新訂マ	1993-013	毛利家文庫史料	山口県文書館	全表示 [隠定表示]
17	新訂マ	1993-015	毛利家文庫史料	山口県文書館	全表示 [隠定表示]
18	新訂マ	1993-017	川谷市立中央図書館蔵種コレクション	川谷市立中央図書館	全表示 [隠定表示]
19	新訂マ	1993-125	毛利家文書	財团法人毛利博物館	全表示 [隠定表示]
20	新訂マ	1993-129	毛利家文庫史料	山口県文書館	全表示 [隠定表示]

図 4. Hi-CAT Plus 検索結果一覧表示

1/6/14 『書目データ』 【書目ID】 00000001 【刊】 採訪マイクロ 【請求記号】 1992-001 【書名】 長崎県立長崎図書館所蔵文書 【原蔵者】 長崎県立長崎図書館 【イメージ】	2 【史料名】 長崎 謹役場絵図 式輪 【利用制限】 10 【細目NO】 8 【文書名】 謹役場絵図 上巻 (長崎会所) 【イメージ】	3 【史料名】 (江戸幕府老中奉書) 【和暦年月日】 (宝永元年)5月9日 【利用制限】 10 【細目NO】 1 【文書名】 長崎奉行江戸幕府老中奉書(原御院様二十三回御法事に御八日参詣につき) 【和暦年月日】 (宝永元年)5月9日 【差出】 長崎奉行(松島正五)・松島馬一(松元義綱)・小佐渡守(小笠原兵重)・土相模守(土相正直)・阿直守(阿直正義)・石尾阿波守(石尾氏信) 【宛所】 【細目NO】 2 【文書名】 長崎奉行江戸幕府老中奉書(近年唐人共作法不宜ため、在留の唐人召集め御書(1)准急度申聞すべきに付) 【和暦年月日】 (正徳4年)5月16日 【差出】 久大新介(久曾義之)・内藤後守(内藤正義)・井河内守(井上正守)・秋田馬守(秋田元矩)・駒木根門(駒木根門)・久松信重(信重) 【宛所】
--	--	---

図 5. Hi-CAT Plus 検索結果詳細表示



図 6. ビューア画面

編纂所外からの利用はできない。加えて、画像ファイルの閲覧、コピー（ダウンロード）、画面キャプチャ、印刷などについて個別の権限制御をおこない、さらには閲覧室の端末ではUSB機器を使用できない設定とするなどのセキュリティ・コントロールを導入している。

6. Hi-CAT との連携

Hi-CAT Plus でのキーワード検索時、既存の『史料編纂所所蔵史料目録データベース』（Hi-CAT）と横断検索ができるようになっている。Hi-CATにおいては「架・番・号」という史料編纂所における配架情報に基づく構造によってデータを管理している。これと Hi-CAT Plus の構造を対応させることで、両者の横断検索を実現した。

たとえば Hi-CAT の「架」に対して Hi-CAT Plus の「UpperDir」を、「番」に対しては「Directory」をあて、Hi-CAT における配架情報

と撮影媒体メタデータであるアーカイブ内の格納位置情報を対応させている。また、「書名」には「書名」を、「冊次表記」には「史料名」をといった具合に、史料の分類階層も適宜割り当てた。

Hi-CAT と Hi-CAT Plus の横断検索によって、採訪史料画像と既存の史料編纂所歴史情報処理システム（SHIPS）データベースとの連携・デジタル史料画像群の参照が可能となった。まだ完了していない採訪マイクロフィルムのデジタル化と一点目録の整備をさらに進めることで、史料編纂所の蓄積してきた史料全体をカバーすることができるようになる。

7. Hi-CAT Plus の活用

前節まで、採訪マイクロフィルムのデジタル化史料の検索・閲覧システムである Hi-CAT Plus について、メタデータの作成と構造化、ビューア、既存システム（Hi-CAT）との連携について

述べた。本節では、この Hi-CAT Plus を活用することで、史料編纂所の業務あるいは日本史研究にどのように役立つか、その展望を記す。

7.1. 研究環境の変化

7.1.1. 画像 1 点ごと検索・閲覧の効用

Hi-CAT Plus によって、史料画像が 1 点ごとに検索・閲覧できるようになることで、これまで活字史料（集）を利用して進められてきた日本史研究環境は大きく変化することになろう。たとえば、ある戦国大名が発給した書状を検索し、その画像をモニタ上に並べることで、容易に各書状の筆跡や花押を比較することが可能になる。

従来は、文書 1 点ごとの検索は困難で、かつその対象がごく限られており、仮にリストが作成できても、マイクロフィルム等から画像（プリント）を 1 点ずつ集めるという極めて手間と時間のかかる作業が必要であった。作業の時間を節約し、研究に費やすことのできる時間を増やす効果を認めることができよう。

もちろん、キーワードによって文書 1 点単位で検索することで、未知の史料と出会うことのできる意義はいうまでもない。たとえばキーワードに年（月日）を指定することにより、特に史料編纂所でおこなっているような編年史料集を作成する際に大変便利である。本来、採訪マイクロフィルムに撮影した文書の内容については、史料編纂所内で十分に情報を共有しておかなくてはならないものであるが、数も膨大であり、それが十二分ではなかった。Hi-CAT Plus によって改めて採訪内容を共有し、編纂に活かすことができる。

7.1.2. 他 DB との連携

現在は前節に述べたとおり Hi-CAT と連携しているのだが、将来的には他の SHIPS データベースとも、さまざまに連携が可能となる。これは Hi-CAT Plus が画像の URL を介して呼び出される機能を実装しているためである。URL の記述の仕方により、1 ファイルのみ、またはある範囲を指定して、Hi-CAT Plus で画像を表示することができる。

たとえば「大日本史料総合データベース」の検索結果からは、現在は「大日本史料」の版面画像の閲覧が可能である。もし Hi-CAT Plus に、そのもとになっている史料の画像があれば併せて表示するようにもできるであろう。

あるいはテキスト系のデータベースから、あるテキストに関連のある Hi-CAT Plus の画像を呼び出すことも考えられよう。

また、「古文書ユニオンカタログ」と連携すれば、刊本や影写本等の情報とともに、Hi-CAT Plus で原本画像を呼び出すという使い方も考えられる。

原本から影写本、マイクロフィルム、写真帳、刊本、テキストデータまで、Hi-CAT Plus がさまざまなデータベースへの波及性を持つシステムであることがわかる。

7.2. 史料クラウド

7.2.1. タイムマシーン

史料は時にその所蔵者を変え、地域を変えて伝えられてゆく。あるいは天災や事故などにより失われることもある。

第 2 節にも述べたように、史料編纂所は、マイクロフィルムを使ったものに限っても、既に 60 年以上の採訪活動をおこなっている。採訪によって、史料編纂所には収集された史料のマイクロフィルムが、採訪年代や所蔵者の情報などとともに残る。史料そのものや所蔵先に変化があっても、史料編纂所には採訪時の史料の様子が変わらず残されているのである。

ある地域の歴史を調べようと思った時に、Hi-CAT Plus を調べることで、当該地域のある年代時点での史料のありようをある程度復元できる、そのように活用することができると考えている。

7.2.2. 所蔵機関への画像提供

Hi-CAT Plus で検索・閲覧できる画像の大部分は、史料編纂所以外の個人や寺社、機関の所蔵する史料のものである。既述のようなセキュリティ・コントロールを導入しているため、たとえば地方の文書館や博物館等の所蔵史料について、その所蔵機関からのみアクセスして画像表示をおこなう、ということも可能である。所蔵機関にとっても、自前のシステムを用意する必要がなく、メリットは大きい。実際にある機関と実現に向けた協議を進行中である。

7.3. 歴史オントロジーの構想

7.3.1. 分散した史料群を再現

たとえば広義の「東寺文書」は、狭義の「東寺文書」、「教王護国寺文書」、京都府所蔵「東寺百合文書」などからなり、これらはいづれももともと東寺にあった史料群である。狭義の「東寺文書」は現在も東寺に所蔵されており、中世以来東寺にとって重要とされてきた文書である。また、「教王護国寺文書」は京都大学が所蔵しており、京都府所蔵「東寺百合文書」のほか、ほかに流出した東寺の文書は少なくない。

このように、もと所蔵先を一つにする史料群であっても、現在は分散して複数の所蔵先に存在している例がある。こうしたものを持てば、「東寺」というキーワードで検索、一覧できるような機能を考えている。それによって、分散する以前の本来の史料群のありようを再現できることになる。

なお、現在の Hi-CAT Plus への登録データは、あくまでも採訪時のものである。史料群が分割

されないまでも、所蔵機関の統廃合や史料群の移動によって、所蔵先名称が変わることもしばしばある。これについても対応することが必要であると考えている。

7.3.2. 関連する史料群どうしの関係を示す

Hi-CAT Plus では、史料群ごとにコードを付して管理することを計画している。これは、史料群どうしの関係を構造化し、コードを用いて表現するためである。

たとえば函館と長崎は、地域としては大変離れたところにあるが、ともに幕末の開港地だったという同じ歴史を持っている。つまり、幕末の函館（当時の表記は箱館）に関する史料群と、同じく長崎に関する史料群については、「幕末開港地の史料群」という共通性を見出すことができよう。したがって、「幕末開港地の史料群」という概念のもとに、「幕末箱館の史料群」と「幕末長崎の史料群」がぶらさがる構造を作ることが可能である。

このように、その存在形態の研究から何らかの連関性が認められた史料群どうしの関係を構造化し、コードを用いて表現するための研究を進めている。

これら史料群の関係は、研究者にとっては明らかなことであることも多いのだが、その概念の相互関係について記述できる仕組みを用意しておくことが重要である。なおかつ、Hi-CAT Plus では常に画像を参照しながらそれらを表現することが可能である。

7.4. ボーンデジタルへの対応

近年の採訪はマイクロフィルムではなく、デジタルカメラによる採訪（「デジタル採訪」）が中心となっており、そうした撮影媒体を持たない収集史料画像へも、同様のメタデータ構造を適用することで、本システムに組み込んでいくことが可能である。これは階層構造化された史料メタデータを持つ本システムの特徴もある。

採訪マイクロフィルムのデジタル化画像の場合、その管理単位（Directory に対応）はマイクロフィルム 1 リールであったが、ボーンデジタルの場合、そのような物理単位がない。そこで、採訪の際に管理のために発行される「採訪コード」を Directory に対応させることとした。このことによって、デジタル採訪画像をアーカイブに格納、さらに Hi-CAT Plus で検索・閲覧することが可能となった。実際に運用をはじめたところである。

8. おわりに

本論文では、本年度運用を開始したマイクロフィルムをデジタル化した史料画像についての検索・閲覧システム「Hi-CAT Plus」について述

べた。第 7 節にさまざま述べたように、いろいろな活用・発展が期待できると考えている。

課題としては、史料編纂所以外の個人や寺社、機関所蔵の史料の画像を中心とするので、その画像に対する所蔵者の権利が重いことから、現在のところ史料編纂所閲覧室でしか利用していただけないという点が大きい。しかしこれは、閲覧室に来ていただきさえすれば、マイクロフィルムの 1 コマ 1 コマを自在に検索して画面に表示できるようになった、ともいえる。さらに、所蔵者の理解を得られたものについては、インターネットを経由した利用の道も閉ざされているわけではなく、今後は順次そうした方向に進んでゆくであろう。

さらに、蓄積した史料画像ファイルと他のさまざまな情報（テキスト、人物、時空間情報など）をつなぐ機能の開発も視野に入れながら、今後も引き続き歴史知識情報の蓄積と検索・提供手法についての研究・提案をおこなってゆく予定である。

付記

本研究の一部は、科学研究費補助金・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」（20222001）および同（A）「ボーンデジタル画像管理システムの確立に基づく歴史史料情報の高度化と構造転換の研究」（23240031）による。

参考文献

- [1] 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、Vol.25, No.1, pp.5-10, 2010.
- [2] 山田ほか：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論文集 人文工学の可能性～異分野融合による「実質化」の方法～, pp.145-150, 2010
- [3] 山田、大内ほか：デジタル史料写真帳：収集史料のデジタル化と検索・閲覧システム、画像電子学会年次大会予稿, 2011