

東京大学史料編纂所における特別推進研究 「前近代日本の史料遺産プロジェクト」 について

横山 伊徳

東京大学史料編纂所近世史料部門
yokoyama@hi.u-tokyo.ac.jp

本稿は、現在石上英一氏を中心に進められている「前近代日本の史料遺産プロジェクト」の概要を紹介し、あわせてそのプロジェクトの成果として浮かび上がってきた歴史学・史料研究における新しいシステムについての展望を示すことを目的とする。そのため、まず東京大学史料編纂所における電子計算機導入の歴史的位置を踏まえ、本プロジェクトの2つの戦略的目標を紹介する。ついで、プロジェクトを支えるシステムの概要を述べ、さらにプロジェクトを構成する4つのグループの位置づけ、目標と到達点を展開して、歴史学のための近未来システムを展望したい。

◎「前近代日本の史料遺産プロジェクト」(JMP) のめざすもの

日本史史料のためのファクトファインディングシステム

本プロジェクトの母体である東京大学史料編纂所は、1895年帝国大学文科大学に史料編纂掛を設置したことに始まる。その前身である史誌編纂掛の中心人物の一人だった久米邦武が「神道は祭天の古俗」という論文によって非職となるという、いわゆる久米事件(1892年)によって、同掛が廃止されたあとの再出発であった。したがってこの種の事件の再発を防止するため、新設の史料編纂掛は掛員に厳重な規約遵守を求めた。この規約の中に、編纂中の史料は「一切他に漏洩するを得ず」という一文があり、久米事件の歴史学への影響を示すことがらとして史学史上知られている。

その後戦後改革があり、また大学闘争を踏まえた大学改革議論の過程で、史料編纂所自身も変化し、より積極的に編纂用収集史料を学界に供する体制を整えてきた。こうした流れの中で、1984年以来取り組んできた電子計算機導入による史料集編纂と収集史料研究のためのデータ蓄積活用という成果を、より早くより直接的に学界に還元したいと考え、90年代に入ってからインターネットに対応する研究と開発を進めて現在に至っている。この過程を通じて、史料分析の1つの方法として電子計算機を最大限活かす、歴史情報研究のスタイルを確立してきた(1984-SHIPS/1997-SHIPS for Internet/2001-SHIPS@BroadBand)。

本プロジェクトは、この計算機に支えられた歴史情報研究・編纂システムをできるだけ普遍化し、日本全国あまねく存在する日本史史料のために、高速ネットワークを活かして内外の研究者・諸機関が協働するためのシステムを目指すものである。できるだけ多くの日本史史料をカバーしてデータマイニング(歴史学的な表現ではファクトファインディング)を行い、そこから導かれた史料や事実、知識の照合・蓄積を可能とするネットワークシステムである。

2つの戦略目標

上記システムを実現するためには、2つの戦略を設定する必要があると考えた。第1の戦略は、史料そのものの属性を記述したメタ・データとそれを蓄積するデータベースに、歴史学の研究上必要とされる現物のイメージデータをリンクした、メタ+イメージリンクの蓄積である。おそらくリンク技術はネットワーク技術としては、目新しい点はないであろう。むしろ、メタ・データとイメージをリンクする原則をどう確立するか、という運用面での技法が問題となる。私は、ここではペーパーベースのドキュメント管理の手法をデジタルデータの管理にも応用する、という考え方をとった。すなわちペーパードキュメントの管理コー

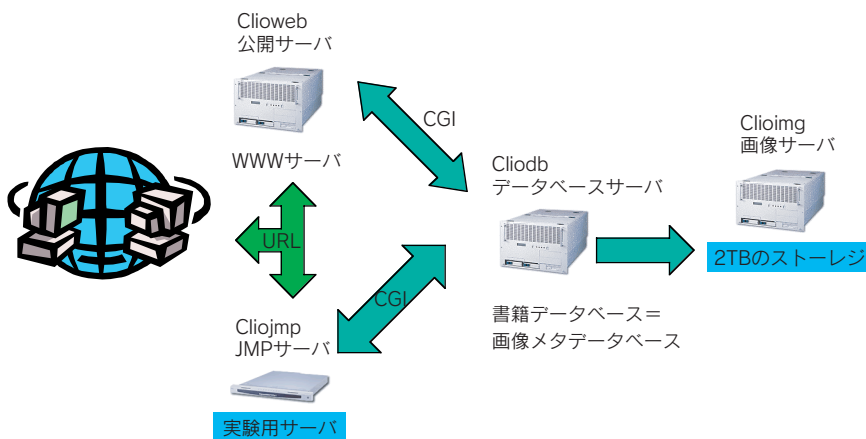


図-1
JMPシステム構成

ドとそのペーパーをデジタル化したデータ（画像，テキスト）の管理コードとを一致させる，という手法である．これにより，ペーパードキュメントの管理データベース（すなわちOPAC）が，そのままデジタルドキュメントの管理データベースに転化する．この管理データベース配下のさまざまなシステム開発がプロジェクト前半の課題であった．

第2は，上記のシステムが対象外としたドキュメント，つまり自ら管理するコードの対象外のデータを参照するための戦略である．これらがハイパーテキストやPDFとなっていれば，一般的な巡回検索エンジンで参照することは可能であろう．しかし，組織による史料管理を前提と考えた場合には，そうしたドキュメントはまた，他組織の管理データベースの下にあるケースを想定するのが妥当であろう．つまり，これら自他複数の管理データベースをネットワークから共通に利用できるようにするコラボレーションが，ポイントとなる．残念ながら，これは技術的にはいくつか課題を抱えている．この解決が，プロジェクト後半の1つの課題となろう．この点は後に触れる．

つまり，すべての史料群がデジタルとなり，そのデジタルドキュメントを共通のしかけで検索できれば，ネットワーク上の分散と協調により，日本列島の歴史史料を覆ってデータマイニングを行うことができる．

◎ JMPシステムの現状と研究グループについて

JMPシステムの概要と特徴

東京大学史料編纂所では現在，SHIPS@BroadBandと呼ばれる歴史情報処理システムが稼働している．これは，ごく大雑把に言えば，WANサービス用にWeb ServerとDatabase Serverを配置し，LANサービスとしてFile Share Serverを配し，これらを1000BASEで接続しているものである．またSHIPSへのアクセス数は，月平均79万件を数えている．

JMPシステムは，このSHIPSを前提に，後に述べるようにそれぞれの目的に合わせ，双方向・事件史・個人史・史料体の4グループによって開発，運用が行われている．JMPシステムのハードウェア上の実体は，CLIOIMGと呼ばれるStorage ServerとCLIOJMPと呼ばれるWWW Serverである．CLIOIMG上の史料画像は，JMPで開発されたImage Display Moduleによって，Database Serverに取められたメタ・データとリンクしており，画像データの各種データベースへの入力やファイルシェアに際して，これらメタ・データを常に参照できるように仕組みられている．

また，CLIOJMPは，Search Engineを搭載し，日本史史料サイトへの巡回検索を可能にしている．このサーチエンジンに使われている形態素解析には，後に述べる日本史語彙辞書が使われる予定である．さらに，このCLIOJMPには，高解像度イメージをWebサービスするためのエンジンも搭載してある（ただし別プロジェクトによる）．

JMPのプロジェクト進行の現状（主として第一戦略）

JMP研究グループ	双方向グループ	事件史グループ	個人史グループ	史料体研究グループ
研究課題	全体システム設計 語彙・画像蓄積	歴史的文脈の計 算機上の表現と その蓄積方法	歴史上の人物の個 人情報 系図・系譜情報の システム化	古文書字体情報 史料の物理的情報 の電子計算機によ る分析・蓄積
構築システム (予定も含む)	画像ストレージ 語彙システム 協働検索	大日本史料総合 システム 古代中世フルテ キスト	人物典拠システム	電子くずし字字典

表-1
JMPの研究グループ

は、史料編纂所の所蔵する史料について、その所在 location を示すメタ・データの蓄積を中心に、作成主体 (個人・組織) についてのオーソリティコントロール、史料分析手段やツール (文字や語彙) のシステム開発、史料に内在する歴史的コンテキストの統合を実現するシステム開発を行っている、という段階にある。これらの JMP システムの特徴は、次の2点にあると考える。

第1は、JMP システムが画像+メタ・データを常に前提に組み立てられていることである。たとえば、文字サンプルのある文書のイメージから抽出する際に、その文書のメタ・データを常に背後に伴って抽出している。したがって、その文字イメージは、どの史料群 (すなわち地域などを特定可能) に属し、誰が、いつごろ作成した文書中の文字であるか、という情報を荷札のように付随させている。これにより、書体の時代的変遷や、地域的偏差などを意識して、文字書体のオーソリティデータを作成することができる。

第2は、近代歴史学なり、あるいはそれを支えてきた史料編纂所100年の資源を活かす、ということである。たとえば、個人のオーソリティデータを蓄積する場合、それをゼロから作るのではなく、『公卿補任』(国史大系本)『柳宮補任』(本所刊行物)のような歴史学が常に道具としてきた典拠を基礎にするということであり、そしてそれに新たに見出された知識を付加するシステムが、JMP システムである。

以上のシステムを稼働する本プロジェクトは、史料編纂所を中心に、歴史研究者、情報学者・計算機科学者から構成されているが、これらのプロジェクトメンバーは、表-1のような4つのグループに分かれている。

JMP 双方向グループ

双方向グループは、ネットワーク資源の構築を第1の課題としている。繰り返しになるが、この資源とは、ペーパーベースのドキュメントの画像とそのメタ・デ

ータである。前者は主として、史料稿本 (『大日本史料』の原稿、完了)、謄写本 (原史料を謄写した写本、進行中)、影写本 (原史料を影写した写本、完了)、本所出版物 (約1,000冊、2001年度分まで完了)、貴重書 (史料編纂所所蔵原本史料) の画像ファイル (267万ファイル) からなり、CLIOIMG に格納されている。一方、これらを管理するメタ・データは、所蔵史料目録データベース (刊本、謄写本、影写本、写真帳、写本、貴重書など史料編纂所が管理する史料図書全体の書誌データを格納したデータベース、ほぼ完成)・編年史料網文データベース (『大日本史料』網文、『史料綜覧』、『史料網文』) を収めたデータベース、ほぼ完成、史料稿本のメタ・データ・古文書目録データベースなどに収められている。各グループや、本プロジェクト以外の研究チームが、史料や版面画像を用いたデータベース作成のために利用している。

双方向グループの第2の課題は、語彙システムの構築である。このシステムの必要性は、まず、インターネットの検索エンジンの精度向上という点に求められる。つまり、日本史史料や文献のための形態素解析に即応した辞書を構築する必要による。この辞書は、現在『古事類苑』索引により、意味体系 (『古事類苑』の編成そのもの) を付与したものとなる予定。さらに、この語彙システムの応用として位置づけられているのが、日本史グロッサリーシステム (仮称) と日本史史料サイトの巡回検索システムである。

双方向グループの第3課題は、本所出版物総合ガイドシステムである。これは、既存の東京大学史料編纂所出版物のデータベースが、1つの書目について1つの機能を実現するデータベースであるのに対して、1つのデータベースで、複数の書目について複数のガイド機能を持つデータベースを開発し、なおかつ、PCレベルのアプリケーションでそれを実現する (特殊なハード、ソフトを要求しない) ことを目指したものである。

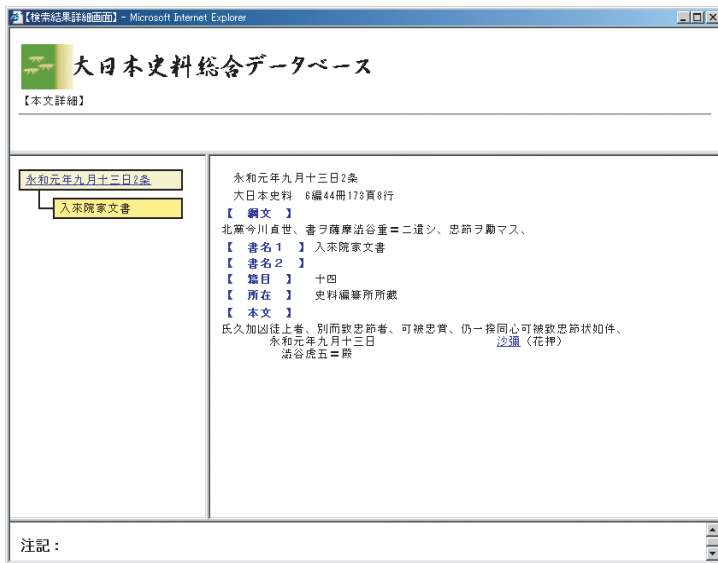


図-2
大日本史料総合システム

双方向グループの第4課題は、史料サイト協働検索システムである。これは2つのシステムを含む。第1のシステムは、日本史史料サイトの巡回検索エンジンの設定である。これはシステム的には導入されており、現在手作業により日本史史料サイトを特定している段階にある。第2のシステムは、日本史史料を所蔵する代表的機関である、歴史民俗博物館、国際日本文化研究センターなどととも、データベース検索を協働して行えるようにするシステムである。これは、本来正攻法では、Z39.50のような共通検索プロトコルを採用して行われるべきである(国土地理院Geographic data clearinghouseを参照)かもしれないが、こうしたシステムは大掛かりでなおかつデータマッピングなどの難しい問題を抱えている。一方、この隘路を解決するために、各機関がデータベース検索のインタフェースを公開して調整するという手法も考えられる。これらの困難は、システム的に未経験であると同時に、データの標準化の未成熟ということが絡んでいると考えられる。

JMP事件史グループ

事件史グループは、歴史上の「事件」のなかから生まれた史料を、史料のコンテキストに即して相互に関連付けるその関連性の蓄積を目的とする。具体的には、第1に『大日本史料』総合システムの開発である。これは、既存の網文データベース・引用書目データベース・索引データベース・編年カードデータベースの各データベースを統合し、それぞれのデータベース内の

レコードと他のデータベース内のレコードのリレーションを形成することで、『大日本史料』という大部でなおかつ複雑な構造を持った史料集のフレームワークをネットワーク上で表現するためのものである。また、このフレームワークに対して、出版物の版面や底本史料画像、花押画像とのリンクを形成することにより、従来の活字による史料集では不可能だった、史料集テキストと原史料・花押の同時表示を可能にした。さらに、『大日本史料』の主たる原稿である史料網文・史料稿本を取り込むことにより、稿本(原稿)データや既刊史料のデータを参照しながら大日本史料編纂を行うことを可能にすることを狙っている。

同グループの第2の課題は、通時代的編年フルテキストデータベースの形成である。奈良時代は『大日本古文书』編年文書、平安時代は『平安遺文』未収文書、鎌倉時代は『鎌倉遺文』のフルテキストを入力することにより、奈良～鎌倉までの通時代的なフルテキストデータベースを構築することを目指している。

第3の課題は、近世史編年史料網文データベースの形成である。江戸時代初期までをカバーする『大日本史料』に比較して、それ以降明治維新までの近世期の史料を対象とする場合は、網羅的なフルテキストデータベースの構築は現実性も利用度も低いと考えられる。そこで、主として『史料網文』、加賀藩史料、池田光政日記、会津藩家世実紀などに基づき、政治史を対象とした改めて近世史編年史料網文を蓄積するものである。

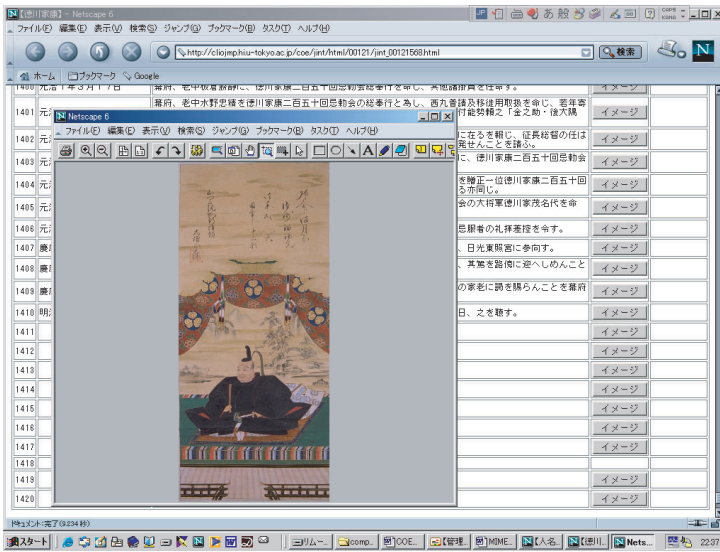


図-3
人物典拠モジュールからその人物の肖像を呼び出したところ

JMP個人史グループ

個人史グループは、第1に、総合的な歴史人物データベースの構築を目的とする。つまり、過去のさまざまな人びとが史料の上に残した諸々の痕跡を個人単位で網羅し、人物情報史料として再構成することを目指す。具体的には、既存の各種SHIPSデータベース、花押や筆跡、肖像画、系図・系譜史料を統合して、有名無名を問わず日本前近代史上人物の情報を名寄せして、総合的に管理・検索する人物典拠モジュールを作成している。システム的には、SHIPSデータベース(網文、肖像、各種索引)から人名と年次データを抽出、人物単位に年次で並び替えたhtmlファイルを作成・蓄積した上に、これらのファイルに対して双方向グループで構築した検索エンジンを巡回させることにより、検索を実行している。

個人史グループは、第2に、基幹の人物誌史料のデータ蓄積にも取り組んでいる。具体的には、「大日本史料」6・7編、公卿補任(古代～近世)、柳菴補任(近世)、忘形見、花押彙纂などを対象としている。これらにより蓄積された人物データを前期の典拠モジュールのデータとマージして同モジュールを拡大していくことを考えている。現状の到達点は、37,000人分の履歴データを作成したが、SHIPSデータだけで約300万人の履歴が形成されると予想されている。

第3に、1人の人物に1つのhtmlファイルを対応させ、ファイルとファイルの関係を記述することにより、人と人との関係＝血縁を表現し、これをつなぎ合わせて、

系図のモデルを開発する研究も行っている。

第4に、個人を特定する重要な指標である花押の数値的分析の研究も行っている。

JMP史料体グループ

史料体グループは、電子くずし字字典の作成と原本史料の物的解析を2つの柱として研究を進めている。第1の電子くずし字字典とは、代表的な草書字典である元禄9年刊『草露貫珠』をベースに漢字と筆記体イメージとの関連づけを蓄積したものである。さらに、史料画像とメタ・データの関連づけの済んだ史料から、書体を抜き出して漢字と関連づけることも可能にした。これにより、切り出した文字に元の史料のメタ・データを付与することが可能となった。これにより文字書体は、時系列データや、作成主体の情報を帯びることになる。これによって、年次による書体変化や地域的書体偏差の追跡を可能とするものである。

第2の原本史料の物的解析のテーマは、上杉文書、大友文書、立花文書などを中心に、紙質、封式、折式などまでも解明できるreproductionの作成の可能性を、DVや高解像度デジタルカメラによって探るものである。

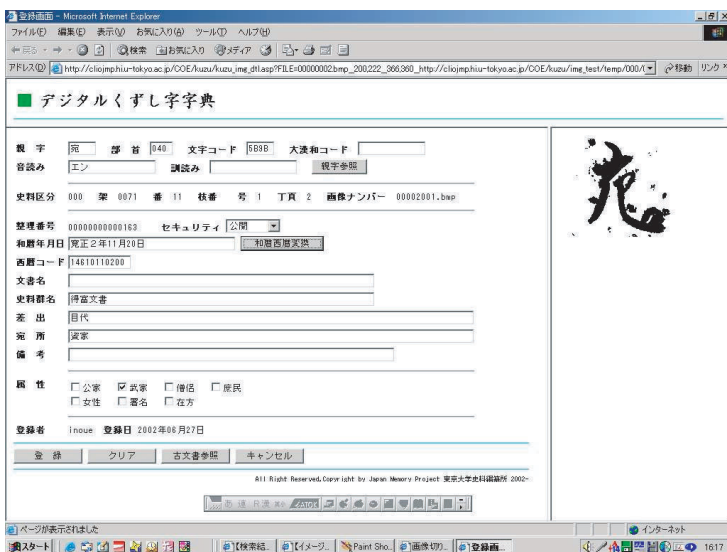


図-4
電子くずし字字典の画面

◎2つの歴史資源をつなぎあわせる

最後に、マネジメントの観点から電子化プロジェクト・モデルとして本プロジェクトを考えておきたい。本プロジェクトは、ちょうど今から100年前活版印刷による本格的史料集を出版したときと同じく、歴史研究のための知識・情報とその媒体との関係を考えるという役割を担っている。そして、100年前との違いは、長い伝統を持つ本のかたちの史料集の長所を最大限いかしつつ、それとは異なる特徴を出す、というところにある。それは、検索システムであったり、画像表示であったり、複製の容易さであったりするであろう。私は、それは2つの全体性を捉えることにありと考える、そのためにシステム設計と運用を心掛けてきた。

第1の全体性は、史料集の基となった原史料の全体性の再現である。特定の史料の有名な文書だけを電子化するのではなく、原史料の全体性を提示し、その中で編纂者・研究者が選択し史料集に載せられた史料がどう位置付くのか、を明示することである。これは、単なる典拠注を超え、原史料の全体的理解へと利用者を導く。第2の全体性は、史料集編纂の研究過程全体の提示である。史料集を作ればすぐに分かるが、手書き文字を翻字すれば史料集となるわけではない。史料を集め、整理し、複製を作り、翻字し、関連事項や関連史料を調査研究し、原稿を書き、校正して史料集はでき上がってくる。このすべての段階で、見出された有用な知識や情報を蓄積して、刊行史料集とならぶ

今様「集解」を構築するのである。

こうしたシステムとその運用にあたっては、史料をめぐる権利関係が複雑な陰を落としている。もっとも、これは電子化プロジェクトではある程度は必然的な問題であり、個人的には、研究情報の自由な流通という立場に立って楽観している。

本プロジェクトは、大きくいえば歴史資源historical resourcesの電子化プロジェクトである。しかし、それはデジタル画像を巨大に蓄積したdigital archivesを目指すものではない。歴史学は、史料に基づく科学である以上、その研究対象である「史料」を蓄積することは大前提である。その史料を活字として出版した版面画像やさらにはその活字メディアを入力した電子テキストを、そのもととなる史料画像とリンクさせ、最後にメタ・データによりそれら全体を管理する。こうしたかたちでのデータ蓄積と管理の業務モデルを作り上げ、日本史史料サイトの1つのモデルとして位置づけたいと考えている。現在検討が進んでいる協働検索が稼働し始めて、多様なモデルによるサイトを編み上げれば、日本史史料の巨大な資源に対して、計算機とネットワークの力を活かしたマイニングを行えるようになる、と確信している。

私はこうした仕掛けを、GILSになぞらえてHILS (Historians' Information Locator Service) と命名し、次のように述べた。「HILSは、拡散する歴史研究のコミュニティへの検索エンジンです。したがって、句切処理に必要な語彙集corporaは、歴史研究の特定のコミュニ

ティではなく全体に対応する必要が生まれてきます。たとえば、『古事類苑』索引など歴史学にとって有用と考えられる語彙集を集積する、などという作業が、歴史研究者のためのインフォメーション・ロケータ・サービスには不可欠となるでしょう。」

この意味するところは以下の通りである。研究は研究対象＝史料のみで遂行できるわけではない。研究に用いる手段＝道具が不可欠であり、これらもまた長い研究の蓄積の中で1つの資源を形成している。史料を読み、解釈し、歴史的文脈の中に位置づけるための資源を情報化していくことが、史料そのものの情報化とともに求められるゆえんである。くずし字からなる史料を読むための電子くずし字字典、さらに文字と文字の組合せによって表現される言葉を蓄積するための日本史語彙辞書・人名典拠システム、そしてそうした言葉によって記述される文に時間軸を与えるための歴史的イベントを詳細な日付によって蓄積する電子年表、などの諸システム構築とデータ蓄積が、プロジェクトのもう1つの狙いである。これらの各システムの基本は、これまでの近代歴史学や史料編纂所の研究蓄積を活かして、

それぞれ紹介したような基幹的史料集の電子化によって形作られている。そしてデータベースであるからこそ、その基幹的部分に、先の蓄積管理されたデータベースから新たに見出されたデータを追加・更新できる。そのような作業は、各コミュニティにおける研究手段の向上にも結びついたものであると考えられよう。

21世紀の日本史史料研究は、蓄積された知識とコミュニティを超えた知識の共有を電子計算機とネットワークを通じて実現していくであろう。私はその研究に従事する者として、コンピュータとネットワークの研究者・技術者とともに、本プロジェクトがその実現に向けた大きな一歩となることを確信している。

参考文献

- 1) 横山伊徳: 東京大学史料編纂所におけるデータベースの構築について 史料編纂・公開機関のデータベース構築の一例として, 歴史と地理, 448号 (1992).
- 2) 横山伊徳: 東京大学史料編纂所データベース SHIPS for Internet, 東京大学史料編纂所報, 第31号 (Mar. 1998).
- 3) 横山伊徳: インターネット時代の歴史学, 歴史評論, No.578 (June 1998).
- 4) 横山伊徳: 公文書館システムについて, 人文学と情報処理, 第23号 (Aug. 1999).
- 5) 横山伊徳: 歴史学にとってのマルチメディア, 人文学と情報処理, 第30号 (Mar. 2001).

(平成14年8月4日受付)