

史学とコンピュータ I

——「日本荘園データベース」の作製を通じて——

福田 豊彦

はじめに

1987年に国立歴史民俗博物館に勤務した私は、以前に東京工業大学にいたという単純な理由によって、歴史研究部の情報委員に選ばれた。東京工業大学に勤務していた時には、ワープロとしてパソコンを使った事はあるが、データベースは初めてで、まさに60の手習いであった。しかし本年3月の定年退職時には、厚かましくも情報委員長として、歴博の情報政策を担当するようになっていた。

もとよりコンピュータの原理も知らない素人であるから、資料課情報係をはじめ、多くの方々のご援助によって、ようやく職責を果たすことができたのであるが、他にも歴博在勤中には、館の情報検索等経費や共同研究経費⁽¹⁾、及び科学研究費補助金⁽²⁾をうけて、「日本荘園データベース」や「吾妻鏡 全文データベース」「玉葉 全文データベース」「製鉄遺跡データベース」などの作製にもかかわっていた。そしてその内の「日本荘園データベース」は、1993年12月には歴博公開データベースの一つとしてオンラインによる公開が実現し、来年3月の科研研究終了を待って、CD版の頒布計画も具体化しようとしている⁽³⁾。

本稿では、これらのデータベースの現状を説明しながら、史学へのコンピュータ利用にかかわる問題点を考えてみたい。これらのデータベースの作製にあたっては、試行錯誤も少なくなかったが、意図的に挑戦した新しい試みもある。この経験が今後多くなるであろう歴史関係データベースの作製に、参考になることもあるのではないか、と思われたためである。

なお本稿では、歴博が目指している考古学や民俗学との協業によって建設しようとしている「広義の歴史学」ではなく、文字的な資料によって研究される「狭義の歴史学」の分野へのコンピュータの利用を主たる課題とする。このため本稿では、そのような狭義の歴史学とその資料を「史学」「史料」と呼んで、広義の「歴史学」と区別することにした。狭義の史学関係のデータベースでは、文字検索にかかわる問題を避けて通ることができず、また文字史料は多くの場合に、3次元空間の性格を持つものとして扱わなくともよい、等々の特殊事情があるためである。

従って本稿では、私が作製に関係した上記データベースのうち、「製鉄遺跡データベース」は検討対象から除いた。実はこのデータベースでは、文字情報を中心としながらも、成分分析値の

表示に計算式を導入したのであるが、そうした試みも、むしろ考古学的資料という特殊性に基づくもので、本稿の対象としては適当ではないと考えたためである。

註

- (1) 平成元～3年度歴博共同研究『荘園データベース構築の基礎的研究』。この研究に関しては、本年度中に刊行される予定の『歴博年報』第1号に、簡単に要約している。なお、後註(3)の第1論文を参照されたい。
- (2) 総合研究(A)『日本荘園データベースの作製と利用に関する研究』、課題番号 04301043〔研究代表者；平成4年度＝福田豊彦・平成5年度＝石井進〕。なお本稿は、この研究報告の一部として提出する。
- (3) 「荘園データベース」に関しては、以前に、「日本荘園データベースへの招待」(『歴博研究報告』第35号・1991年)、及び「荘園と地名―日本荘園データベースへの招待―」(『歴史と地理』第442号・山川出版社・1992年)を書いている。また、本年春の歴博企画展『荘園絵図展』に試作中のCD版を公開したが、その宣伝を兼ね、「データベースで見る日本の荘園」(『電楽』1993-6号)として紹介していただいた。併せてご参看いただければ幸いである。なお、このデータベース作製の時間的経過の報告は、歴博研究報告に収録された前稿に記したので、本稿では省略する。

1 史学とコンピュータ

急速なコンピュータの発達と普及によって、コンピュータの史学への利用も試みられ、道は次第に開かれている。

我々の祖父の時代には、論文は毛筆で書かれ、父の世代にはペンを使い、我々の時代には万年筆を使うことが認められるようになって、その訂正も、書き直し→切り貼り→インク消しへと変化して、「便利な世の中になったものだ」と、半ば皮肉をこめてうらやましがられたものである。しかしコピー機器の普及によって、鉛筆と消ゴムの使用が可能となり、原稿用紙の誤字の切り貼りも遠い過去の思い出となった。そして最近では、論文やレポートにもワープロが広く使われている。ワープロは原稿修正がたやすくできる事が特長で、私のように悪筆・悪文の徒には救いの神であるが、その修正は痕跡も残さない完璧さである。まさに急激な文房具革命ともいえようが、義務教育段階でコンピュータをマスターしてくる学生が、創造的な学問分野に活躍する20年後には、歴史教育や史学研究の分野にも、日常的にコンピュータが利用されるようになることは疑いないであろう。

しかし文房具としては、算盤、鉛筆、クレヨン、整理用カード、パンチカードから簡易辞書まで、多様な機能をもつコンピュータは、使用目的を明確にしなければ、その作品であるデータベースは、実用に耐えない混沌に陥ることが予想される。歴博開館当初に進められた「館蔵資料データベース」に関しては、学問的な立場からする率直な批判がなされず、経験が共通の遺産になっていないことが残念であるが、これは、何にでも使える汎用性を目指したがために膨大なものとなり、速度は遅く、却って中途半端なものになってしまったのではなからうか。そして更に言えば、個別性を尊重する歴史学と普遍性を基本とする計算機との間には、基本的に相容れない性格があるように思われる。

これまでコンピュータを、多くの機能をもつ文房具のように捉えてきたが、勿論それは、利用者の側からする比喩的な表現であって、正確には電子回路を用いて数値計算・論理計算を行う計算機で、文字の処理に注目すれば、二進表現の採用によって、文字をも電氣的に高速処理することを可能にした計算機ということができよう。従ってコンピュータは、普遍的・恒常的で繰り返すことの可能な、数量化できる事象を扱う事を得意としている。

それに対して史学には、古くから法則性を拒否して一回性を重んじる思想があり、事物を独自ので個人的なものとして把握しようとしてきたところに特徴がある。戦後の科学万能の風潮の中で、熱意をこめて歴史学の学問としての特徴を明らかにし、狭義の科学とならぶ社会的な役割を説いた上原専禄氏は、歴史学を大要次のように定義した（『歴史学の概念』『世界の歴史⑥』毎日新聞社・1954。『歴史学序説』〈大明堂・1958〉に再録）。

歴史学という学問は、生活現実のうちで、研究者の問題意識や生活意識にとって知る値打ちがあると想像されるものを選びとって観察の対象とし、それを個人的なものとして認識し、また、その生活現実を生成の一回性と性格の特殊性においてとらえようとする、理知的探求を内容とする経験科学の一つである。

この上原氏の定義は、当時急速に優れた成果を蓄積しつつあったマルクスやウェーバーの思想を踏まえながらも、18世紀以来の史料批判に基づく知的認識の学問として、科学万能主義への批判をこめ、歴史学の社会的職分を鋭く提起したもので、その視角は当時の歴史研究者に大きな自信を与え、戦後に急成長を遂げた現在の日本史学界の潜在的基礎常識となっている。そしてここで最も熱意をこめて説かれているものが、自然科学に代表される科学的方法との違いであった。

この上原氏の視角は、古代中世の日記記録に富むわが国の史学では、特に受け入れ易く、不可欠のものであろう。余談ながら、本書によって、微積分学の基礎を開いた自然科学者ライプニッツが、『ブラウンシュヴァイク編年録』という歴史書を著していたことを知り、改めて歴史に憧れを抱くようになった者も、私1人ではないのではなかろうか。

このような史学の特質を考慮すれば、いくらコンピュータが発達しようとも、史学のあらゆる分野をコンピュータが席卷する事はないのであって、有効性はかなり限られている。従って我々がコンピュータとつきあう第1歩は、歴史学のどのような分野で、コンピュータのどのような機能を使うのか、という視点を明確にして出発することであろう。

その意味では、差し当たり私が史学の分野で利用しようとしたコンピュータの機能は、いずれも文字検索とその数量的な把握にある。まずは「日本荘園データベース」の内容と現状、作製にともなう諸問題をみ、ついで「全文データベース」に及ぶことにしたい。

2 「日本荘園データベース」の現状

「日本荘園データベース」には、内容としては、「荘園データベース」と「荘園文献目録データ

ベース」の2つを含んでいる。いずれも多くの方々の協力を得て今日の姿になったのであるが、前者は私が、後者は小島道裕氏が、主として担当してきた。従って「荘園文献目録データベース」に関しては、小島道裕氏が報告されるはずのものであるが、今回は同氏の了解を得て、必要な範囲で触れることにした。

「荘園データベース」に登録された荘園には、中世史料にみえる「庄」の他に、御厨・御園・牧・柚・保・別符などを併せて収録しており、近世史料にみえる地域呼称としての庄名も含まれているが、純然たる国衙領である郡・郷は対象としていない。以下では、それらを総合した用語として「荘園」を使い、狭義の「庄園」と区別するが、現在登録されている荘園の総数は約8,800である。また「荘園文献目録データベース」には、個別荘園関係の論文のほかに、一般の荘園関係の論文を含んでいる。従ってその選別の基準や記述の微細な内容には、県別に行われた選択者による多少の差異を避けられなかった。その文献の採録数は、今後も年々増加するであろうが、現在のところは、平成5年末の分までで約3,900である。

また、利用する機器と利用の形態によって区分すると、この「日本荘園データベース」には、パソコン上で動くパソコン版と、歴博ホストコンピュータに載せられたオンライン公開用のホスト版、及び平成5・6年度文部省科学研究費によって作製されたCD版、の3種類があることになる。

この3種類は、データ内容は近似するが、データとして最新の基本になるものはパソコン上に置かれている。以下ではこれを「荘園DB」「荘園文献DB」と呼ぶことにする。パソコン上で動くこの「荘園DB」は、市販のデータベース・ソフトであるdBASE III プラスの上に構築されているが、「荘園文献DB」はZEROの上に作られている。パソコン版・ホスト版・CD版にはそれぞれに特徴があるが、保存上では最も不安のあるパソコンのハードディスクに入っているデータを基本とした理由は、市販ソフトを使用するのでデータの修正や追加・消去を日常的に行うことができ、必要に応じて統合・分割や他のデータベースへのデータ転換も比較的容易で、コンピュータに素人の我々にも管理できること、などによるものである。

歴博ホストコンピュータに納められているホスト版は、電話回線による公開を目的としたもので、そこでは荘園データと文献目録データの両者が統合されている。またCD版は、前述のように科学研究費の助成をうけて作製中のもので、地図と絵図情報を取り込むように工夫したところに特徴があるが、そこでも両者は統合され、ある程度の制約はあるが、相互に乗り込む事ができるようになっている。なお、画面上の表現はそれぞれに工夫があって、そのソフトの作製は、ホスト版はファコムハイタック株式会社に依頼し、CD版は大日本印刷株式会社に依頼している。

このホスト版は、平成4年度に館内試行を行い、平成5年11月の公開予定で作業が進められている(12月にこの公開が実現した)。

(1) 荘園データベースの内容

「荘園データベース」の内容は、次の19項目からなっている。1例としてパソコンデータのプ

リントを表1に掲げた。この「荘園DB」では、パソコン・ディスプレイの3画面にわたって表示される。ホスト版とCD版は、表示形式では相互に多少の違いがあり、データが少ない場合には詰めあわせて表示されているが、基本的には同じ構成をもち、記載内容も同じである。

この各項について、簡単に説明しておこう。表でもわかるように、この「荘園データベース」では、区切符号として「・」又は「。」を使っており、ホスト版とCD版では、この符号で囲まれた単語が検索対象になっている。

なお、各項ごとに、使用する文字の大きさ（全角か半角か）と、総文字数には制限があり、それを越えることは出来ない。最終的には、オンラインによる公開用データの作成を目的としたデータベースでは、技術的にやむを得ない制約である。各項の*印の下に、使用する文字の形式と制限総字数を記しておく。

① [荘園コード] * 半角数字 7文字

このデータベースでは、荘園はそれぞれ7桁の固有のコード番号を持っている。基本的には国・郡に固有のそれぞれ2桁の数字と、その荘園を表す3桁の数字で構成されている。しかし、本来は所属郡不明（未勘郡）を意味した3・4桁に99をもつ荘園の中にも、郡名の記載のあるものが見られるように、研究の進展によって所属郡が明らかになり、本来の国郡番号との間にズレが生じた場合も少なくはない。

② [国名] * 全角漢字 2文字

荘園の所在国を示す。このデータベースでは、原則として古代の行政区画によっている。境界地方では、時代によって国所属が変動している場合もある。

③ [郡名] * 全角漢字 2文字

荘園の所在郡。原則として『和名類聚抄』の国郡区分によっており、中世の郡所属とは一致しないこともある。なお、所属郡不詳の場合には「未勘」としている。

④ [荘園名] * 全角漢字 7文字

荘園の名称。使用漢字は固有名詞であっても、後述のように統一をはかっている。従って例えば、紀伊国阿氏川庄は阿旦河庄、豊前国穉田庄は稗田庄、伊勢国瓜生御厨は瓜生御厨などと表現して、外字の使用を避けた。しかしパソコン版とCD版では、(紀伊国持田庄)・(越中国榎田庄)・(駿河国穠津御厨)・(山城国毘原庄)のように、一部に外字を使用したものがある。

実際の検索では、パソコン版は中間一致が可能であるが、オンライン誘導型及びCD版は、前方一致が原則になっている。従って例えば「黒田」と入力すれば、黒田庄・黒田出作・黒田保・南黒田御厨などが同時に検出される。黒田庄のみを検出したい場合には、「黒田庄」と入力する(黒田荘ではない)。また、CD版で庄・保や牧・御厨を包括的に求めたい時には、この荘園名の欄で、「*庄」「*保」「*牧」「*御厨」と入力すればよい。なお、オンラインやCD版では中間一致ができないので、例えば「田中」と入力しても、小田中庄は検出対象から除外される事になる。

⑤ [フリガナ] * 半角カナ 12文字

歴史仮名遣ではなく現代仮名遣による。

⑥ [重複コード] * 半角数字 23文字

例えば、下総国船橋御厨は夏見御厨と呼ばれ、持田庄が笠田庄とも書かれたように、荘園には異称のあるものが少なくない。時にはまた、本庄や新庄、東庄と西庄のように、分割される場合もあった。そのような場合には、例えば船橋御厨と持田庄がそうであるように、一つのカードを基本カードとして記述を集中し、他のカードを参考用の脇カードとして処理することを原則とした。この欄に番号が記載された荘園は、そのカードも参照されるように期待している。

⑦ [参考市町村] * 全角漢字 10文字

荘園の比定地として、現在の自治体名を一つだけ掲げる。荘園の中には、島津庄のように薩隅日3か国にわたる荘園もあれば、畿内の錯圃型荘園のように、1町村に数個の荘園が並立している場合もある。このデータベースでは、比定地として一つの自治体を参考市町村として掲げる。その荘園の範囲がその自治体に限られるという意味でもないし、その自治体の範囲のすべてが中世にはその荘園に含まれていたという意味でもない。

また、重複して登録された荘園では、参考市町村の記載は本カードのみに限り、原則として、その他のカードには市町村名を記入していない。検出数を問題にする場合など、『主要情報欄』をみれば、この欄の記載の有無によって、簡単に重複を避けることができるように工夫している。

また、CD版で地図上に表示する場合には、参考市町村名の記載のない脇カードの分は、地図上には表示されないことになる。従って同じ荘園が地図上に重複して数えられることはない。

⑧ [市町村コード] * 半角数字 5文字

その自治体の市町村コード番号。自治体名とそのコード番号は、原則として平成4年度版によっている。

⑨ [明治村字名] * 全角漢字 370文字

清水正健編『荘園志料』は、荘園の範囲を、通常、明治20年代の市町村の字名によって示している。それは、近世の村名に通じるものが多いという点でも貴重である。歴博では、別に『旧高旧領データベース』が公開されているので、将来の連結を考慮し、この欄に大きなスペースを割いた。

⑩ [史料村字名] * 全角漢字 125文字

中世におけるその荘園内部の地名。郷名には特別な意味のある場合があるため、「某々郷」と表示したが、史料に村名でみえるものには、原則として「村」の字は省略した。

⑪ [領家・本家] * 全角漢字 50文字

荘園の領家・本家。検索の便宜のため、院領荘園や後院領・勅旨牧などには「皇室領」、九条家領や近衛家領には「摂関家領」、延暦寺関係の荘園には「山門領」などの総括的な呼称を併記し、また皇太神宮領には「伊勢神宮領」、その「内宮領」「外宮領」を兼ねた御厨・御園に

は「二宮領」という統一した呼称を併記して、検索の便宜を図った。

⑫ [初見年・和暦] * 全角漢字 7文字

その荘園の史的な初見年を示す。荘園の立荘年ではない。和暦年には、10は十、20は廿、30以上は三十、三十一などと漢字を使用する。

⑬ [初見年・西暦] * 半角数字 4文字

史料初見を示す西暦年で、4桁のアラビア数字。検索の際の時期範囲の指定にはこれを使用することになる。なお、その場合、10世紀以前には0を付けて4桁の数字とする。例えば、10～11世紀初見の荘園を求める時には、「0901～1100」のように指定する。

⑭ [出典] * 全角漢字 75文字

主要な史料の出典。

⑮ [遺文番号] * 半角のカナと数字 150文字

『平安遺文』『鎌倉遺文』等の文書番号。例えば、『平安遺文』一二二五号文書、『鎌倉遺文』二二六五七号文書は、それぞれ「へ1225」「カ23657」のように表示される。ここでは、「ナラ」は『寧楽遺文』(上巻頁数)、「へホ」は『平安遺文 補遺編』、「へキ」は『平安遺文 金石文編』、「へタ」は『平安遺文 題跋編』、「ナ」と「ナン」は『南北朝遺文』、「キヨ」は『経塚遺文』の文書番号を意味し、「シ」は『大日本古文書編年文書』の巻数と頁数を表現している。なお、史料が多く、ここに書ききれない荘園もあるが、それらの多くは有名な荘園であり、研究論文があるので、それらを参照されることを期待している。

⑯ [記録類] * 全角文字 75文字

関係史料が記録類にみえる時、その記録名(『 』内)と日付(元号年、月、日)を示す(数字も全角文字)。ここでも、制限文字数では記載しきれず、日付などを省略したものがあつた。なお、字数節約のためにこのデータベースでは、『吾妻鏡』は『東鑑』と表記している。

⑰ [地名辞典] * 全角文字 75文字

『荘園志料』、『角川日本地名大辞典』(「角川地名」と表示)、『平凡社日本歴史地名大系』(「平凡地名」と表示)、『平凡社大百科辞典』(「平凡百科」と表示)などにより、その荘園の立地や比定地などを記す。なお、「荘園志料=なし」及び「荘園志料=記事なし」とは、それぞれ『荘園志料』に記載のない荘園、及び史料が掲載されていない荘園であり、後者は主に、近世の地名としての「庄」であることを意味する。

⑱ [備考] * 全角文字 150文字

大田文の公田数や、鉄による年貢納入などのような、その荘園についての特徴的な事項をこの欄に記す。絵図のある荘園は「絵図=」として所蔵者等を挙げている。しかしオンラインによる公開データベースでは、この欄は地名辞書の欄と共に検索の対象にならないし、CD版では「絵図」記載のある荘園だけを検索対象に加えた。

⑲ [関係論文] * 二者択一の論理型

その荘園に関する研究文献が、「荘園文献目録データベース」に収録されているか否かを、有(T)・無(N)の別によって示す。

(2) 荘園関係文献目録データベースの内容

「荘園関係文献目録データベース」は、次の11項目からなっている(表2)。いわゆる典型的なカード型データベースで、特に解説の必要もなからう。パソコン版で使っている ZERO は、各項目ごとの制限字数をきめる必要もなく、使用文字も混用できる便利なものである。しかし、そのデータをホスト版や CD 版とする際には固定長を定めなければならない、ホスト版では文字も統一しなければならない。このため、例えば関連荘園は最大 5 荘園で切り捨てることになった。ここに記した制限字数は、その場合の文字数である。このデータベースでは、区切符号に「,」又は「,」が使われている。

表2 「荘園文献目録データベース」の内容

番号	項目名	内 容
1	著 者	小山 靖憲
2	著書・論文名	中世村落の展開と用水・堺相論—高野山領名手荘と粉河寺領
3	出典 1	『中世村落と荘園絵図』(東大出版会)
4	出典 1 年月	1987. 11
5	出典 2	
6	出典 2 年月	
7	国 名	紀伊
8	荘 園 名	名手庄
9	荘 園 コ ー ド	5703001
10	備 考	用水, 堺相論, 絵図

① [論文コード] * 半角数字 4桁

それぞれの論文に与えられた整理番号。登録順に自動的につけられるが、配列は自由に変更することができる。

① [著者名] * 全角文字 15文字

論文の執筆者名、及び著書の編著者名。連名の場合は「,」を入れて連記。刊行物には委員会や団体名の場合もある。

② [文献名称] * 全角文字 50文字

論文名または書名。市町村史には章節名を含む事がある。

③ [出典 I] * 全角文字 125文字

その論文を収録した雑誌。または書名と出版社名。

④ [出典年月 I] * 半角数字 7文字

その論文または著書の刊行年。すべて西暦年により、月とも7桁の数字で示す。例えば、平成3年5月は1991.05と表示。刊行月不明のものは00。

⑤ [論文出典Ⅱ] * 全角文字 125文字
その論文が再録された場合の刊行物名。

⑥ [出典年月Ⅱ] * 半角数字 7文字
再録された刊行物の刊行年月。

⑦ [関連国名] 全角文字 20文字
論文が主として扱った荘園の所属国名。

⑧ [関連荘園名] 全角文字 70文字
論文が主として扱った荘園名。

⑨ 荘園コード 半角数字 70文字

⑨の荘園の荘園コード。ホスト版とCD版は、このコード番号によって、『荘園論文DB』と相互につなげている。

⑩ [論文備考]

「絵図」「領家・本家」「開発」「班田」「村落」など、論文のもつ視点によって、相互に関連づけるためのキーワード。本来は、題名や出典が検索対象とならないホスト版を考慮して設けたもので、その多くは内容によらず、論文名から切り出した。将来はこの欄を充実させ、検索対象として有効に利用されることが期待される。しかし、論文名にあらわれない事項を網羅的に列記する事は、将来においても困難であろう。

(3) 使用文字について

最終的に公開を目的として作成される歴博データベースには、文字規定があり、使用漢字は1983年に改正された日本工業規格(新JIS)第1・第2水準指定文字に限ることになっている。オンラインによる公開を前提として作られたこの荘園データベースでは、外字の使用は避けなければならない、この制約はやむを得ないであろう。文字検索のためには、異体字などについては更に統一をはかって使用を制限するか、或いは異体字相互を結び付けるために、検索の際に使用する歴博検索ツールを開発する必要があるが、今はその問題には触れない。

JIS規格に限ると、荘園に不可欠の文字にも使えないものがあり、それらは借字などに頼ることになる。竿→算、析→料、鋌→廷、役→役、職→職のように、常識的に判断できると思われる文字転換の他に、下記のように頻出する用語もある。

借字一覧	実 例
勝→榜	「四至榜示」
籙→録	「撰録渡荘目録」
轄→幹	「平維幹」
氏→互	「阿互河庄(紀伊国)」
蕪→稗	「稗田庄(豊前)」

瓜→瓜 「瓜生御厨（伊勢）」

しかし、パソコン版とCD版では、紀伊国持田庄の「持」、駿河国藤津御厨の「藤」、越中国栞田庄の「栞」、山城国厩原庄の「厩」など、僅かではあるがやむを得ず、一部に外字を作製したものもある。CD版は外字の組み込みが可能であるためである。ホスト版ではこれらは、「かせ田庄」「かい津御厨」「くぼ田庄」「みかの原庄」等と表現されている。

なおこの荘園データベースでは、重複コードで本カードと脇カードを結び付け、荘園を異称でも検索できるように工夫した。従って、和泉国日根庄・下総国船橋御厨や紀伊国阿弋河庄・越中国宇川庄などは、それぞれ日根野庄・夏見御厨・安世河庄・鶺河庄でも検索ができるようになっている。しかし「河」と「川」、「中」と「仲」、「太」と「大」、「禰」と「祢」、「曾」と「曽」、「富」と「冨」のように、当時から混用されていても、常識的に判断できると思われたものは、煩雑を避けて重複登録をしていない。従ってそのような荘園を検索する場合には、例えば白河庄、太田庄、田仲庄などは、白河・白川、太田・大田、田仲・田中などと併記して検索することが望ましい。なお、固有名詞には「々」の使用を避けているので、佐々木庄などは佐佐木庄として検索する必要がある。

3 「日本荘園データベース」作製の経験から

(1) 作製の経過と残された課題

日本荘園データベースは、試験的に着手してから今日までに5年余を費やした。この間には試行錯誤も多いが、機器の進歩は予想以上に目ざましかった。作製の時間的経緯は前稿に記したが、今後の史学関係データベース作製のため、ごく簡単に経過と問題点を整理しておきたい。

荘園データベースの作製は、歴博館内の田中稔・八重樫純樹・西沢奈津子と私の4名が、1987年度の追加予算に試験的な入力経費を申請したところから始まるが、その段階に立てられていた基本線は、次の3点程度の漠然としたものであった。

- ① 荘園データベースは、最終的にはオンラインによる歴博公開データベースの一つとするが、経費の節約と素人でも動かせるという点に配慮し、パソコン上で作製する。
- ② 作製は、まず清水正健編『荘園志料』収録史料からはじめ、それを『平安遺文』と『鎌倉遺文』の史料によって修正増補する。具体的には、『荘園志料』の必要なデータをマークし、それを入力してデータベースの基本設計を行う。別に、『平安遺文』と『鎌倉遺文』の索引を利用して荘園カードを作り、それによってデータの修正と追加を行う。
- ③ 試験的なデータ入力は、たとえ途中で作業が中断されても、それだけで使えることを考慮し、関東8か国と伊豆・甲斐の10か国とする。その結果をみて、下巻の東国→下巻の西国→上巻の近国→上巻の畿内へと拡大する。最終的には、『荘園志料』の入力に3年とみて、データの修正を含めて公開は5年後になると予想された。

この1987年度の追加予算に試験入力認められたが、手元のメモによると、『莊園志料』の関東分線引きの作業が始まるのは、88年2月のことであった。その時点での入力項目は、「莊園コード」「国名」「郡名」「莊園名」「重複コード」「明治村郷名」「史料村郷名」「領家本家」「史料出典」「史料初見年」の10項目であった。

その入力結果をもとに、上記4名で検討し、「莊園 DB」のシステムを作ることになったが、ここで情報の専門家である八重樫氏から、「フリガナ」と「参考市町村名」「自治体コード」をあげることが強く主張された。フリガナは検索と配列のため、自治体名は他のデータベースと結び付ける将来性に配慮したものである。これに対しては、かなりの技術的困難性と作業量の増加が予想されたため、激論になったが、入れることに決着した。この2項目を追加したために地名辞典類を参照する必要が生まれ、後に「地名辞典」の項も置くことになったのである。

また、近い将来の目標として「莊園関係論文目録」の作製があげられ、「論文」の有無の欄も付けておくことになったし、史料初見と研究上の便宜を理由として、後には「記録類」の欄も作ることになった。

こうして『莊園 DB』は市販データベースである dBASE の上に作られたが、項目としては当初から予定していた「遺文番号」と「備考」欄を含めて前記19項目となり、一つの莊園の表示に3画面を要する大きなデータベースになった（作成は株式会社カンテックに依頼）。また作製の作業としては、全ての莊園について、当時刊行中の角川書店または平凡社の地名辞典を検索する、という大きな仕事加わっていた。そしてまた、1991年に『鎌倉遺文』が完結したのをうけ、『平安遺文』『鎌倉遺文』のデータを収録したほかに、鎌倉時代末期までの金石文や木簡関係の史料も、一応は採録することができた。

その点では、当初予定していた公開時期から1年程度の遅れで完成し、また途中から作業に加わった小島道裕が中心になり、当初は日程に入れていなかった「莊園関係文献目録 DB」も作られて、1993年12月に両者を併合して公開の運びとなったことは、おおよそ順調な進行ということができよう。これも、1989～91年度に行われた共同研究『日本莊園データベース作製の基礎的研究』のメンバーの方々の協力と激励によるところが大きかった。

この間、1992年度には、歴博博物館資料調査委員会の制度を利用して、当初の予定にはなかった府県別のデータ点検も行われ、内容としても一般への公開に耐えるものになったと考える。その委員会では、原則として国ごとに担当者を定め、プリントしたカードに赤を入れていただいた。この中では、データの修正だけでなく、南北朝時代以降に初見の莊園もかなり追加された。また「文献目録データ」の補正もお手伝いいただき、地方史関係を中心に大幅に増強されたのである。

表3に、国ごとの担当者のお名前をあげた。この点検がなければ、庄園名の読みも正確を期し得なかったであろう。入力データの点検作業は、自分の書いた原稿校正の経験しかなかった我々には、予想より遙かに難かしいものであり、単純な入力ミスもここで発見されたものが多かったのである。なお、表の中で*印を付した方は、89・91年度に実施された共同研究において、前記

表3 荘園データベース国別修正担当者

担 当 国	担 当 者	勤 務 先
大 和	*泉 谷 康 夫	奈良教育大学・教育学部
山 城	池 田 好 信	京都府立総合資料館
河内・和泉・摂津	堀 内 和 明	大阪府立鳳高等学校
伊 賀	勝 山 清 次	三重大学・人文学部
伊 勢	*稲 本 紀 昭	京都女子大学・文学部
尾 張・参 河	*稲 葉 伸 道	名古屋大学・文学部
遠江・駿河・伊豆	湯 之 上 隆	静岡大学・人文学部
甲 斐	秋 山 敬	山梨県庁・生活文化課
相 模・武 蔵	福 島 金 治	神奈川県立金沢文庫
武 蔵	田 代 脩	埼玉大学・教養学部
安房・上総・下総	岡 田 清 一	東北福祉大学・社会学科
近 江	高 橋 昌 明	滋賀大学・教育学部
近 江	太 田 浩 司	長浜市立長浜城歴史博物館
美 濃・飛 騨	清 田 善 樹	聖徳学園岐阜教育大学
信 濃	井 原 今朝男	長野県歴史館準備室
上 野	*海 津 一 朗	東京都豊島区史編纂室
下 野	新 川 武 紀	栃木県立宇都宮女子高校
陸奥・出羽・常陸	岡 田 清 一	東北福祉大学・社会学科
若 狭・越 前	佐 藤 圭	福井県立朝倉氏遺跡資料館
加 賀・能 登	棚 橋 光 男	金沢大学・文学部
越 中	楠 瀬 勝	高岡法科大学・法学部
越 後・佐 渡	田 村 裕	新潟大学・教育学部
丹 後・丹 波	黒 川 直 則	京都府立総合資料館
丹波・但馬・播磨	馬 田 綾 子	梅花女子大学
淡 路・播 磨	小 林 基 伸	兵庫県立歴史博物館
因 幡・伯 耆	錦 織 勤	鳥取大学・教育学部
出 雲・石 見	井 上 寛 司	島根大学・法文学部
美作・備前・備中	三 好 基 之	ノートルダム清心女子大学
備 後・安 芸	*坂 本 賞 三	神戸学院大学・人文学部
安 芸・備 後	*岸 田 裕 之	広島大学・文学部
周 防・長 門	田 中 倫 子	山口芸術短期大学
紀 伊	*小 山 靖 憲	和歌山大学・教育学部
阿 波・讃 岐	丸 山 幸 彦	徳島大学・総合科学部
伊 豫・土 佐	川 岡 勉	愛媛大学・教育学部
筑前・豊前・肥前	*有 川 宜 博	北九州市立歴史博物館
肥 前・壱 岐	川 副 義 敦	県立佐賀西高等学校
豊前・豊後・日向	飯 沼 健 司	宇佐風土記丘資料館
肥 後	*工 藤 敬 一	熊本大学・文学部
薩摩・大隅・日向	五 味 克 彦	鹿児島大学・文学部

5人の館内メンバーとともに、中心になってこの研究を推進された方々であるが、92・93年度の科学研究費の研究分担者として、CD版の作製にも参加していただいている。

もとより、ご協力いただいた方々の要望がありながら、実現できなかったことも少なくはない。そのような問題点を掲げ、残された課題として解決の糸口を探ってみよう。

① データの追加……「中世の荘園研究では、国衙領も含めて荘園公領制として捉えることが常

識化しており、保・別符だけでなく郡・郷も対象とした方がよい、」というご意見は、89年の共同研究開始の当初から提起され、他の方からも指摘された点である。

保や牧を対象に入れた理由の第1は、庄園と同様に、主要なものは『莊園志料』に収録されており、『平安遺文』『鎌倉遺文』の地名索引を利用して作ったカードで補充することができたためである。しかし郡・郷にはその条件がない。地名索引の郷名には庄園内の郷が多く、中世に国衙領であった郡・郷名を取り出すには専門的知識が必要であって、学生アルバイト等ではできない事、大田文などから求めるとすれば、史料残存の偶然性による偏差が激しくなり、数的比較の対象にもならない事、などが予想されたため、この提案は将来の課題とした。

この課題では、システム変更は必要なく、データ数の増加だけであるが、問題はデータの収集方法にある。それが今後に残された主要な検討課題であろう。

- ② データ領域の増加……「遺文番号が途中で切れている」「記録類には他にも入れる必要のあるものがあるが、その場所がない」というご意見がある。

データ収集が進むにつれて、当初設定した字数では書ききれない項目が増加した。一部は備考欄への書き込みなどによって処理したが、「遺文番号」「記録類」「史料出典」などにも、データを途中で打ち切った庄園が次第に目だつようになったのである。

この解決にはシステムの設計変更が必要であるが、パソコンの3画面に納めようとした現在の方式では、その余地は殆どない。オンラインによる歴博の公開データベースとしては当面このまま運営されようが、将来『南北朝遺文』データの付加などが課題となった際に、全面的な改訂を望みたい。

なおこのデータの作製は、『莊園志料』を基礎として遺文文書で追加修正する事を基本としたため、記録類からのデータ収集まではとても手が廻らなかった。前述のように、この項は後に付加したもので、兼夷の『玉葉』が一部に『玉海』となったものがあるように、記録名称の統一も充分ではない。公開利用による声を聞きながら、この表記の統一も将来考えていただきたい。「康正二年引付」「康正国役段銭引付」「康正段銭引付」などと表記された「康正二年造内裏段銭并国役引付」のように、史料出典の欄にも統一をはかることが望ましいものがある。

- ③ データ項目の変更……「参考市町村として一つの自治体を選ぶことは正確ではないので、いくつかの自治体名を併記するように変更してはどうか」「史料初見年だけでなく、庄園の立荘年の項目が欲しい」というようなご意見も聞かれた。

そのような情報は「地名辞書」と「備考欄」に記入するようにしているが、オンラインによる公開では検索対象とはならず、データ項目の改編や増設も当面は困難である。当面はまず、備考欄の記述を充実しながら、必要に応じて検索用のキーワードを作るなどの方策も考えられよう。しかし将来、機器の機能が向上し、検索時間の大幅な短縮が可能になれば、パソコン版と同様に中間一致も可能になるのではなからうか。CD版には特にその可能性が高いように思われる。使用者が独自に備考欄に追記し、「莊園データベース」をパーソナルなデータベース

に育てて使用できるような、機器の開発と性能の向上に期待している。

(2) 検索の実際

「日本荘園データベース」を実際に利用すると、どのように操作すれば、どのようなことができるか、実例によって紹介してみよう。

実際の操作はパソコン版とCD版、ホスト版でかなり違うし、ホスト版でも、簡易方式の誘導型とベテラン向けのコマンド方式は全く違う。そして素人の私には、そのコマンド方式は扱えない。従ってここでは、パソコン版を中心に、CD版との違いを考えてみよう。ホスト版の操作方は、資料課情報資料係で作成した手引きを参照していただきたい。

差し当たり問題を検索に限定すると、パソコン版とCD版の相違の第1は、前者が「荘園データ」と「文献目録データ」を分離しているのに対して、後者はそれが結合され、自由に照合できることであり、第2は用語の検索において、前者がワープロのように中間一致をとるのに対して、後者は前方一致を原則とし、必要に応じて後方一致を行うことにある。そして第3には検索速度の違いを挙げねばなるまい。CD版が遥かに速いのである。

その第1の点では、CD版が遥かに便利で、パソコン版でもウィンドウ機能などを利用すればある程度の対応はできようが、第3の使用時間の問題は、機器の進歩をまたなければならぬ。第2の点では、CD版は地名辞書や備考欄を原則として検索対象にできない事になる。しかし、例えば「一牧」や「一保」を対象にしようとするする場合、『荘園名』において、パソコン版で「牧」「保」と入力すれば、「某々牧」「某々保」の他に、「牧庄」「保科庄」等が同時に紛れ込んでしまうのに対して、CD版で「*牧」「*保」と入力すれば、それらは除外される。どちらが便利とも簡単には言えないであろう。そして以上の点に関しては、ホスト版はCD版と原則的に同じである。

具体的な検索の実例を示す事にしよう。

〔**荘園の名称**〕……『荘園名』を選び、「黒田」と入力したとする。

パソコン版では、次の20荘園が検出される。

伊賀＝黒田庄，同庄出作，黒田新庄，同本庄。伊勢＝黒田庄，黒田御厨(2)，南黒田御厨(2)，北黒田御厨，黒田御園，黒田牧御園。尾張＝黒田庄。近江＝黒田庄，黒田江西庄。加賀＝黒田新保。越中＝黒田保(2)。播磨＝黒田庄(2)。

CD版では前方一致を原則とするので、その内の3荘園が減り、17荘となる。「黒田庄」と入力すれば、パソコン版でもCD版でも同じ7荘園が表示される。どれが良いかは簡単には言えず、ソフトの性格を知って使いこなす必要がある事になる。

この黒田の事例でも、全国に同一名称の荘園がかなり多いことがわかるであろう。このような荘園の名称では、「庄」「保」の名称には「田」のつくものが多く、牧に「野」や「島」が多いことと対比して、興味深い問題である。中間一致が可能なパソコン版では、そのようなことも確認

できるが、CD版やホスト版ではかなり難しいであろう。

[初期荘園の分布]……9世紀以前のいわゆる「初期荘園」の分布を調べる。

『庄園名』=パソコンでは「庄」、CD版では「*庄」と入力。

『領家・本家』=「東大寺、西大寺」と入力。

『初見年(西暦)』=0701~0900と指定。

8・9世紀に初見の荘園は、東大寺領と西大寺領を合わせて79庄、その内3庄は重複コードによって重複して数えている事がわかるので、それを除くと76庄である。その分布は五畿内に12庄、近畿地方全域を加えても24庄であるが、越前~越後の北陸地方には37庄と、半数がここに分布し、いわゆる「北陸型」の荘園分布が明瞭に認められる。

この分布は、CD版で地図上に表示すれば、いっそう明瞭に見てとれるのであって、恐らく歴史研究の初心者への教育的効果は絶大であろう。

なおホスト版では、1項目に複数の検索はできないので、東大寺領と西大寺領を別々に検索しなければならない。

[皇室領荘園の時期的変遷]……最大の荘園領主である皇室の所領は、時期的にどのように増加するであろうか。史料初見年は荘園の成立年ではないが、数が多いので、大略の趨勢をつかむ事はできよう。

①『領家・本家』=「皇室」と入力。

『初見年(西暦)』=0701~1000と指定。

8~10世紀に初見の皇室領荘園は60、重複1を除くと59荘であるが、このうち「庄」は18、「牧」が41、「勅旨田」が1つ認められる。ここには『延喜式](このデータベースでは延長五年とした、但しこの数には兵部省の官牧は含まれていない)の出現という偶然的な要素もあるが、初期の皇室経済における勅旨牧の大きな比重は認めてよからう。

②『領家・本家』=「皇室」。

『初見年(西暦)』=1001~1100。

11世紀初見の皇室領は40、重複を除くと39。内訳は「庄」36、「牧」2、「御厨」1である。

③『領家・本家』=「皇室」。

『初見年(西暦)』=1101~1200。

12世紀初見の荘園は439、重複を除いても401を数え、内容は「庄」353、「牧」6、その他42である。特に鳥羽院政以降の爆発的な荘園の増加を確認する事ができる。そしてこの時期には、関東・東北から九州の末端まで、皇室領荘園が全国に分布している。初期荘園と比較して、12世紀に盛行するこの寄進型荘園の分布状況の特長は、CD版の地図でいっそう明瞭に知る事ができよう。そこでは重複登録も自動的に排除されている。

④『領家・本家』=「皇室」。

『初見年(西暦)』=1201~1300。

13世紀初見の荘園は251, 重複を除くと222。内訳は「庄」199, その他 23 である。鎌倉期の史料の増加もさる事ながら, 荘園制はなお最盛期にあったように思える。

⑤『領家・本家』=「皇室」。

『初見年(西暦)』=1301~1400

14世紀初見の皇室領は160, 重複を除くと154で, 増加率は明らかに減少している。その中で内訳が, 「庄」123, 「牧」1 に対し, その他の形態の所領が 30 もみえることは興味深く, 荘園の衰退とともに, 新たな経済源を求めた貴族層の姿がここにかがえるのではなからうか。そして15世紀になると, 初見荘園は25(内その他5)と急減し, 畿内近国への集中度が高まり, 16世紀には初見荘園を1件も検出できなくなっている。

このような皇室領のあり方は, 数が多いので, 摂関家領や中央の権門寺社領など, 他の諸荘園の趨勢とも一致する事が多い。しかし12・13世紀に急増する伊勢神宮の御厨の分布は, 明らかに東国に集中しており, その中で数少ない西国の御厨の中に年貢として鉄を納めるような特殊な荘園がある事や, 摂関家に寄進された奥州藤原氏の荘園が, 何れも「内国」といわれる国司支配下であって, いわゆる奥六郡や山北三郡のような根拠地には一庄も認められない事など, 荘園においても数的把握によって研究の糸口となりそうな事項は少なくない。

そしてそこでも, CD版の地図表示には迫力があり, 教育の場において特に威力を発揮しそうに思われる。

(3) 多様な公開への道=機種と特徴

パソコンで動かされる「荘園 DB」のシステムは, 市販ソフトの上に作られた 0 SYOUEN. PRG によって管理されている。それは試作段階の後にも, 出力形式などに幾度かの微修正を加え, また最初は国毎のフロッピーディスクによる運用を基本としていたものを, 現在ではハードディスクに置くように作り替えている。この作製と修正は株式会社カンテックに依頼したが, その費用は, 大型コンピュータのソフト作製費と比較すると, 桁違いに低廉な価格である。また, 新史料の出現によってデータの修正・追加・除去を行うことは, 予想以上に頻繁な日常的作業であった。新しく開発された別の市販ソフトを使った『文献目録 DB』の場合も, それは全く同じである。

もしこのデータベースが, 初めから歴博ホストコンピュータの上に作られていたとしたならば, 期限内の実現は到底不可能であったろう。その意味では, 基本をパソコンにおいて市販データベースソフトを利用する, という当初計画は正しかったことになる。

オンラインによる公開を目指すにしても, CDの作製を目標にするにしても, “完成された刊行物をそのまま利用するのではなく, 新たにデータを収集して手作りのデータベースを作製しようとする場合には, パソコンを利用した方が経済的にも時間的にも遥かに有利である”ということが, 「日本荘園データベース」作製の経験から得られた第1の結論である。

ともあれ「日本荘園データベース」は、パソコン版、CD版、ホスト版の3種が作られた。それぞれに使用目的が違い、それぞれに特徴があるが、もう一步立ち入って使い勝手などを比較し、今後の参考に供することにしよう。

言うまでもなくコンピュータは器械であり、器械では「慣れ」が第1である。従って使い勝手も簡単には比較できない。その意味では、私が最もよく使ってきたものがパソコン版であり、CD版は作製の過程にあり、ホスト版は横からみているだけ、という不公平な立場にあることも事実である。以下の記述は、あくまでコンピュータに素人の、利用者の立場からの一つの経験に過ぎない事は、お断りしておかねばなるまい。

パソコン版を作製した第1の目的は、オンラインによる歴博公開データベースの基礎的作業として、そのデータを調整するとともに、これを使いながら、現実の歴史研究に使えるような形態を考えることにあった。このためには、データの追加・修正・除去が簡単にでき、同時に共同研究として使用実験が行えるものでなければならなかった。パソコンとして普及し、市販ソフトも多い NEC の PC9801 を前提とし、比較的普及率の高い dBASE III の上に 0SYOUE.N. PRG を作製することにしたのはこのためである。

当時歴博にあって基本に据えた機器は PC9801-VM、歴博の他にあるものも同機種であり、自宅のものは F2 であった。データが集まるとハードディスクが必要になり、移動の容易さとデータの増加を見越して、緑電子のポケディを利用する事にしたが、この段階の機器では、検索時間の制約を考慮しなければならなかった。

則ち、修正などに使ってみた結果、使い勝手としては1件の検索所要時間を40秒以内におかないと不便で、それにはデータ数は1,000件程度で区切る必要があったのである。こうしてデータは、① TOUGOKU (東海道は遠江以東、東山道は信濃以東、北陸道は越後・佐渡の17ヶ国、現在1,166荘園)、② HOKURIKU (近江・美濃・飛騨と越中以西北陸道の8ヶ国、1,290荘)、③ TOUKAI (伊賀・伊勢・志摩・尾張・参河の5ヶ国、1,369荘)、④ KINAI (畿内5ヶ国、1,662荘)、⑤ SANIN (山陰道8ヶ国、835荘)、⑥ SANYO (山陽道8ヶ国、1,032荘)、⑦ NANKAI (南海道8ヶ国、658荘)、⑧ SAIKAI (西海道11ヶ国、753荘)と、全国を8区分し、別に備考欄や村郷名欄を除いた簡便型の全国の抄録版 ZENKOKU を作ることにした。やや不自然な地域区分とも受け取られようが、これは、国によって荘園数にかなりの差異があり、伊勢や大和は1国で700荘を超えるという事情もあって、多分に便宜的な区分である。

現在も私の手元にあるものは、そのように分割したままで使っているが、全国版の作り替えにも結構の手間がかかる。もし5年前に一般の市場にでている機器が、同じ PC9801 でも現在のよような 486 型であったならば、全国を分割せずに一本にしても2分程度で検出できるであろうし、大型のラムディスクを使用すればかなり時間は短縮できるであろうから、その方が良かったかも知れない。しかしそうした機器の性能の問題は、市販ソフトの選択などにも言える事で、余り意味がないであろう。歴博共同研究でこのデータベースを試用しておられる方の中には、他の市販

ソフトで動かしている方もあるし、データの結合や分割は素人にもできる事である。ここではむしろ慣れが重要で、使いこんだソフトを使用の方が便利なのである。

前述のようにパソコン版は、本来はホスト版作製の準備作業であったが、使ってみるとそれ以上の意味がある。検索において文章中の字句を自由に検索できるため、CD版やオンライン版では検索対象ではない地名辞書や備考欄が検索できること、及び利用者が必要に応じてデータを付加したり書き換えたりできることである。荘園データベースの備考欄には、現在でも絵図のある荘園については、その名称や所蔵者を記入してあるし、年貢として鉄のような特殊なものを納める荘園にはそれを記述している。しかしこの欄は、ホスト版では、検索対象となっていない。しかしパソコン版であれば、この欄を選んで「絵図」或いは「鉄」と記入すれば、そのような荘園を簡単に選び出す事ができる（CD版では「絵図」を検索できるようにした）。つまりパソコン版では、研究者がこの備考欄に必要と思われる事項を書き加えるならば、それも検索できる事になる。パソコン版の「荘園DB」は、容易にパーソナルな荘園データベースに変身するのである。

このようなパーソナルなデータベース作成という点では、「荘園文献目録DB」において、いっそう威力を発揮するに違いない。データだけでも20メガを超え、フロッピーで18枚になる「荘園データ」と、データだけならばフロッピー1枚に入る「荘園文献目録データ」では、パソコン版としての普及に大きな違いがでるだろう。もし「荘園データ」をパソコン版として普及させようとするならば、フロッピー2枚程度に納まる抄録版にするか、全国を8地区位に分けた地域版を作る必要があるのではなかろうか。それでもハードディスクが普及した現在では、オンライン方式より簡単な公開方法かもしれないし、このようなデータベースが研究者に公開されるならば、新しい文献の補充や、備考欄への切り出しに利用するキーワードの書き込みなどが日常的に行われ、研究に利用されるようになるであろう。

時間は多少遅くとも、小回りがきいてデータの修正も簡単で、システムの修正も自分で出来るし、他の使い馴れたデータベースへの転換も可能なパソコン版は、素人にとって最高である。これは何よりも、研究者が自分で必要とする情報を備考欄に付加し、自由に検索することが出来るという点で、歴史研究者向けの、パーソナルなデータベースである。しかしここには、「著作権」を如何に保護するか、という大きな問題がある。その点でやはり、差し当たりはCD版が大きな期待の星であろう。

しかもこれは、検索結果を地図上に表示し、目でみて結構楽しめる。しかしデータの更新はパソコンほど簡単には行えない。そのようなCD版の性格を考えると、歴史教育の場面で利用するものとしてはこれが最も適しているだろう。今春の企画展『荘園絵図展』にお目見得させたように、博物館の展示なども、これを基本にして加工するのがよいのではなかろうか。

もちろんそれは研究に役立たないという事ではない。例えば、もし歴博で作製されている「貿易陶磁文献目録データベース」や「城館城郭データベース」のような同時代のデータが、同じCDに組み込まれ、同じ地図の上に表示できるようになるならば、思いがけない歴史学上の新発

見があるのではなからうか。そのように複数のデータベースを開発して同じ地図画面に表示する事は、全く同じシステムが利用できることで技術的に容易であるし、歴博の目指す新しい歴史学の建設にも貢献すると思われる。

初めての歴博作成のCDとなったこの「日本荘園データベース」CD版は、2年間の科学研究費で作られたもので、経費としてもさほど大きくなかったが、「荘園データベース」と「荘園文献データベース」の結合を実現し、荘園絵図もある程度は取り込んで、予想以上に楽しいものにすることができた。しかしそれでも、500メガの容量を持つCDには隙間が多いのである。

このようなパソコン版やCD版と比較すると、素人にとってホスト版は、不器用で経費も余計かかり、速さ以外は魅力に乏しいように思われる。この「日本荘園データベース」ホスト版では、素人のために誘導型システムを開発していただいたが、複数の事項にわたる検索は、1次検索・2次検索と重ねていく事になるため意外と煩わしく、速度が速いという実感も余りわかない。なお、検索の実例で記したように、現在のシステムでは、同じ項目に同時に2つ以上の語句を検索できないため、種々の不便が予想される。技術的な難易度を判断する能力は筆者にないが、そのためのソフト改訂は、遠からずぜひ実現していただきたいと考えている。

以上のような感想は、あくまで素人のたわごとで、コンピュータ本職の方々からは別の指摘があるに相違ない。しかし少なくとも現在では、大型コンピュータを動かせる本職が作成し利用するデータベースではなく、史学という文字の世界でも、必要性に応じて開発し、手軽にデータを作り替えながら利用するという、多様化の時代になっていることは確かであろう。

次にみる全文データベースは、本質的に現在のオンラインによる公開は困難であって、更に多様な公開方式開発の契機ともなると思われる。

しかし残念なことに、すでに予定の紙数を使い果たしてしまった。全文データベースをめぐる諸問題は、現状報告をも含めて別の機会に譲ることにしたい。

追記

本文中では、パソコン上での『荘園DB』の修正について、「優れた現在の機器を使えば、全国のデータを1本化しても2分程度で処理できるであろう」と述べた(394頁)。その後、全国8,900のデータをつなげて1本化したところ、検索などにも4分を超え、カード除去などの作業には、時に数十分を要することが判明した。全国を8地区に分けて進めた『荘園DB』の修正作業は、現在の優れた機器を使っても、現実的で妥当なものであった。データの数量増加と検索などに要する処理時間の関係は、等差級数的な対応関係ではなく、数量が少ないうちは殆ど意識されないが、ある所から等比級数的に時間がかかるようになり、間もなく実用に耐えなくなって、システムの根本的な変更が必要になる、というようなものであるらしい。僅かな経験を普遍化する事は危険ではあるが、今後試みられるであろう歴史史料のデータベース作業の参考として、ここに記しておく。

(元国立歴史民俗博物館歴史研究部)