Rekishow Authoring Tool の可能性
― 歴史研究のためのデータベースとの関連を中心に ―

平山　勉（ひらやま　つとむ）
（慶應義塾大学グローバルセキュリティ研究所研究員）
友部　謙一（ともべ　けんいち）
（経済学部教授）

はじめに

われわれのプロジェクトでは，2002 年度より，Rekishow Authoring Tool（以下，Rekishow と略）の開発とそれを使った危機管理研究を進めている。Rekishow は，「暦象」という時間属性と地理属性を持ったデータにもとづいて分析を進めるが，特に 2002～2003 年度にかけては，数量分析のための機能開発を進めてきた。つまり，「いつ」「どこで」「どれくらい」という 3 つの属性を持った歴史事象を分析の対象としている。

具体的に言えば，年々集計されている，府県別の伝染病統計や都市別の物価データなどを使って，分析を進めようとしている。ところが，この種のデジタル・データを入手することが，困難であることが分かってきた。ネット上に，さまざまなデータベースが公開されているが，歴史研究にとって有効なデータベースはそれほど多くはない。

本稿では，こうした現状をふまえながら，Rekishowの開発と利用から明らかになった，歴史研究のためのデータベースについてまとめてみたい。

1．データベースの 3 類型

90 年代初めの三田新図書館には，カード目録がまだあった。その一方で，すでに OPAC も設置されており，1986 年以降の図書は検索することもできた。とりわけ，キーワードが含まれるケースを全てピック・アップしてくれる「中間検索」には，とても助けられた記憶がある。アナログなカード目録よりも進んだ世界が，OPAC にはあるように感じられた。その後，検索の方法も豊かになり，多くの人々にとって，OPACは身近なデータベースの１つになっている。

OPAC のような検索を目的とした目録型データベースは，最近になって，より進化してきたように思う。たとえば，朝日新聞と読売新聞の戦前紙面については，PC で読むことができるようになった。検索エンジン[1]の結果から，ダイレクトに紙面へと移動して閲覧するシステムは，実に便利である。マイクロフィルムや縮刷版を，疲れた眼を酷使しながら 1 枚ずつめくっていた状況とは格段の違いだ。このようなスムーズな閲覧を目的とした画像型データベースは，今後，ますます普及するように思う。

その一方で，社会経済史の分野には，長期経済統計データベース[2]のように，専門性の高いデータベースがある。このデータベースでは，経済史の研究に直結する統計データを，Web 上で入手することができる。その意味では，分析のための素材型データベースと言えるだろう。このように，歴史研究を支えるデータベースには，目録型・画像型・素材型の 3 つのタイプがある。

2．歴史研究のためのデータベースの要件

目録型と画像型のデータベースは，主に研究環境の向上に貢献する。これに対して，素材型のデータベースは研究内容そのものに直結してくる[3]。引き出されたデータの加工などを通じて，分析は深まっていくものだ。しかし，素材型のデータベースは公開されることが少なく，同時に，作成者以外の人が使うことも少ない。不幸なことであるが，歴史研究では素材型データベースは，広く共有されるには至っていない。さらに，素材型データベースがどのような性質を備えるべきかについても，継続的に議論なされているわけでもない[4]。そこで，素材型データベースの要件について考えてみたい。

最初の要件は「拡張性」を有することである。新しい歴史資料が発見される可能性は常にある。明日になれば，思いもよらぬ資料が発見されて，新しい

データを得ることができるかもしれない。そうした状況の中で，データベースを新規に作り直すのではなく，幸運なハプニングを順次吸収することが，研究の進展には必要になる。

次に，さまざまなデータを組み合わせることで，新しい知見を得られるような「発見性」を備えていることである。研究の本質は資料や論理のオリジナリティの追求にあるのだから，誰が使っても同じような分析結果しか出ないのであれば，素材型データベースとして魅力的とはいえない。使い手によって，新しい分析や解釈が可能となるような構造を持つことが大切だ。

また，データベースの構築過程を確認できるような「透明性」を維持する必要もある。実証主義的な歴史研究にとって，データ出所の開示は必要不可欠だ。データの内容が正確かどうかを確かめたいという欲求は，史料批判的な意味でも常にある。加工されたデータが無批判的に一人歩きするような状況は，是非とも避けたい。そして，最後に，データベースを構築したり，その過程で発生するミスをチェックしたりするコストを，できるかぎり抑えるような「簡便性」を備えていたい。

3．カード・タイプ・データベース

①拡張性　このような要件を満たすデータベースは，カード・タイプ・データベースになるだろう。たとえば，内務省衛生局による伝染病統計について考えてみよう。図 1 にあるように，府県，調査年，病名，罹患 or 死亡，数値，単位といった項目の他に，出所情報として，表題，収録誌なども項目として入力していくことで，カード・タイプ・データベースは構築されていく。言い換えれば，ひとつひとつの数値に対して，調査年・調査地(府県)・出典などの属性情報が付されていることになる[5]。

図　1

ここで例に挙げた『衛生局年報』のように，表の形式が統一されているデータを扱う場合には，表組みを決めて，表計算ソフトのスプレッド・シートに入力していく方が効率的かもしれない。しかし,『衛生局年報』に記載された数値は，各府県から報告された数値を内務省がまとめたものであり，その報告には期限が設けられている。一方で，府県レベルの統計書には，期限後に把握されたものを含んだ数値が記載されていることが少なくない。つまり，内務省と府県ともに，同じ年，同じ地域(府県)，同じ病気を対象としていながら，発表される数値に違いが出るのである。

スプレッド・シートに入力していった場合，こうした状況に対応するには，おそらく注を付けることになるだろうが，その作業は極めて煩雑である。だからと言って，その段階で正しい数値を選択するという行為は，回避されなければならない。その理由はいずれの数値も，歴史的事実を反映しているからである。2 つの数値を選択・加工する方法は，研究者の判断に委ねられるべきであろう。ここに拡張性を備える意義がある。

②発見性　カード・タイプ・データベースの特徴のひとつは，いわゆるリレーショナル機能にある。一般的には，顧客情報と注文伝票との連結などを例に，効率的な情報管理のためのツールとして，リレーショナル機能は解説されている。しかし，こうした予定調和的な使い方に，リレーショナル機能を限定する必要はない。

たとえば，上述の府県別の伝染病統計のほかに，その予防法教育を目的とした街頭映画について，日時・会場・参加者数・映画タイトルなどが記された，衛生当局の活動報告書を見つけたとしよう。ある府県では全く上映記録がないかもしれないし，逆に，大がかりな啓蒙活動があったことが分かる府県もあるだろう。そうした情報の有無も含めて，この活動報告書をカード・タイプ・データベースにして，伝染病統計のデータベースと連結することもできる。その結果，街頭映画の内容（タイトル）と伝染病患者数との間に，新たな関係性を発見できるかもしれない。こうした発見の可能性は，歴史研究を進めていく上で，何よりも必要とされる。こうした

発見性を備えることで，単なるデータファイルではなく，データベースとしての価値が出てくるのである。

③循環性　拡張性と発見性とが同時に備わって，素材型データベースは価値のあるものになっていく。これらの要件に，カード・タイプ・データベースは，もうひとつ重要な性質を与えてくれる。それは,「循環性」である。

素材型データベースからのデータを使って研究をする場合,実際に採用する分析ソフトは様々である。ある人は，既存の統計パッケージソフトを使うであろうし，またある人は，それとは性格の異なるGISソフトを使うであろう。そして，採用するソフトによって，要求されるデータ・フォーマットもさまざまである。

カード・タイプ・データベースは，この要求に応えるクロス集計[6]機能を備えている。1つの数値に1枚のカードを割り振ることで，つまり，データを最小単位で蓄積することによって，表組みを自在に変えることが可能となっている。その結果，さまざまな分析ソフトに対応するデータ・フォーマットを作成することができる。おそらく，これから開発されるさまざまな分析ソフトにも，対応することが可能であろう。

4．Rekishow Authoring Tool

「暦象オーサリング・ツールによる危機管理研究」プロジェクトでは，4つの要件を備えたデータベースを構築して，具体的な分析を行っている。蓄積されているデータを挙げれば，①上述の官庁統計などの数量データの他に，②小作慣行調査などのテキスト・データ，③近代日本農業技術年表などのイベント・データ，④研究論文などで発表されたデータなどである。いずれも,カード・スタイル・データベースとして蓄積している。これらの統一的なデータベースにもとづいて，プロジェクトではRekishowを使った分析を深めている。

ここで，簡単にRekishowについて紹介してみよう。Rekishowは，図2にあるように，X軸に経度，Y軸に緯度，Z軸に時間を配置した時空間を再現している。ユーザーは，この時空間を航行することによって，歴史事象を関連づけていき，HistoryのAuthorとなる。図2は，ある伝染病の毎年の府県別患者数を表示しており，個々のオブジェクトの集合が日本の形をしているのが分かるだろう。この画面は，同時に4つまで表示可能であるから，さまざまなデータ系列とあわせて認識することで，分析がより豊かになっていく。また，個々のオブジェクトは，時間，場所，数値，出所情報や，関連するテキスト・データやイベント・データを，詳細情報カードで表示することができる。

Rekishowそのものは，もっと汎用性を高める方向で開発が進められている。汎用性の含意は，使われずに眠っているデータベースのリサイクルと，それによるリスクの緩和である。

活用されないデータベース（データファイルと言った方が正確である）をリサイクルするために，RekishowではデータをCSVファイルでインポートするようにしてある。個々のユーザーは，自分自身がかつて使っていたデータを再利用することで，Rekishowによる分析が可能である。

しかし，自分で用意したデータだけでは，十分な分析ができないこともある。データ件数そのものが不足していることもあれば，別種のデータを追加・拡張する必要がある場合もあるだろう。いずれにしても，それなりのコストが要求される。Rekishowでは，ネット上のサーバーに個々人のデータを蓄積して，それらを共有することで，こうしたリスクを緩和するシステムを採用している[7]。

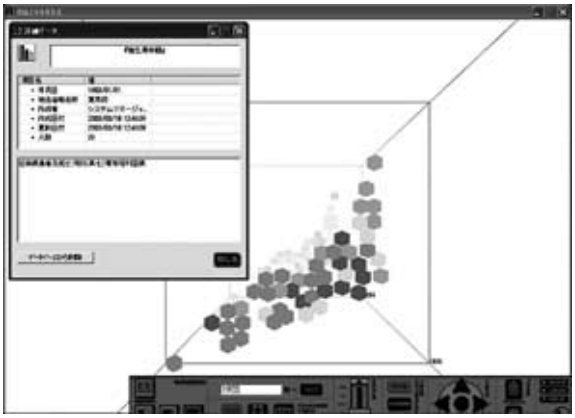

図　2

おわりに

最後に，Rekishowと素材型データベースとがも

たらす可能性についてまとめてみたい。

ひとつは，データのリサイクルを通じて，分断されたデータベースを統合することである。Rekishowによる分析を通じて，さまざまなデータが，サーバーに蓄積されていくことになる。単に，データベースのみを統合するのではなく，実際に分析をする中で起こるニーズにもとづいて統合しようとすることで，作成者の間で「共有意識」が醸成されるだろう。

次に，リンクされたデータによって，抽象化や概念化を抑制した研究が展開されるようになる。これまでのように，分断されたデータベースで研究する場合，データベースごとに「結論」が出され，それらを止揚してきたように思う。しかし，データベースを連結させることによって，データは個別具体的に豊かになっていく。Rekishowがデータごとに付けた詳細情報は，この豊かさを実現している。

最後に，IT技術の限定的な利用を克服することである。これまでのデータベースは，検索と閲覧を目的とするものが多かった。しかしながら，Rekishowと素材型データベースとによって，今後は，分析を目的としたデータベースの利用を，ユーザーは志向するようになるだろう。

暦象オーサリング・ツールによる危機管理研究
（文部科学省 学術創成研究）
http://www.fcronos.gsec.keio.ac.jp/home.html

注

1）確かに，『「朝日新聞戦前紙面データベース」ユーザーズマニュアル』（昭和10年〜20年編，2002年）からは，記事編集を60名以上の人が担当していることが分かり，記事内容のキーワードの統一性について疑問も残る。しかし，見出しについては全文テキスト化がなされているため，データベースの構築過程で発生する恣意性の問題は回避され，一次資料としての性質も保持されている。

2）一橋大学経済研究所による長期経済統計データベースは，集計プロセスについての透明性が低く，素材型データベースとしては十分なものとは言えない。

3）仮に，新聞記事がテキスト化されていれば，構文分析などさまざまな分析が可能となる。しかし，戦前紙面のテキスト化を，商業ベースで実現することについては，議論の余地が残るだろう。なお，神戸高商（現神戸大学）による経済・経営関係の新聞記事切抜のデータベース（画像・全文テキスト）については，次のURLを参照のこと。http://www.lib.kobe-u.ac.jp/sinbun/index.html（2004.6.30参照）

4）数少ない議論を伝えるものとして，尾高煌之助・松田芳郎編（1983）がある。

5）カード・タイプ・データベースに直接入力する場合，表題や収録誌などのように，同じ内容の続く項目をひとつひとつ入力することは，極めて煩雑な作業となるとともに，入力エラーを確認する手間も膨大なものとなる。この簡便性と透明性の問題については，平山　勉（2002）を参照のこと。なお，紙幅の都合により，以下では拡張性と発見性に限定して論じる。

6）さまざまなソフトがクロス集計機能を備えているが，総じて，表頭と表側に指定できる項目が1つであることが多い。そのような中で，2001年versionで開発が終わってしまったLotus Approachが，最大4項目まで同時に指定できる点を，改めて指摘しておきたい。

7）第三者の作成したデータを，安心して使うためには，出典情報が明記されている必要がある。カード・スタイル・データベースとして蓄積することで，この問題を解決している点を確認していただきたい。

参考文献

・尾高煌之助・松田芳郎編（1983），『日本経済統計データベース編成の課題と方法―シンポジァムの記録―』統計資料シリーズ26，一橋大学経済研究所・日本経済統計文献センター
・平山　勉（2002）『近代日本の時系列統計におけるメタデータの構造とデータベースの構築―内務省「衛生局年報」を事例として―』KEIO-GSEC Project on Frontier CRONOS；Working Paper Series No. 02-010
・永島　剛・市川智生（2003），『「衛生局年報」における法定伝染病・種痘統計とその入力：2002年度F-CRONOS疾病班作業報告（1）』KEIO-GSEC Project on Frontier CRONOS, Working Paper Series No. 02-015