

人文学研究におけるデータ中心アプローチの可能性

－『斐太後風土記』データベースを事例として－

松森 智彦

データ中心アプローチ（DOA）とは、情報工学のシステム開発分野の用語である。データベースを中心にシステムを設計・構築する手法で、今日広く用いられている。この手法を借用して人文学研究に持ち込み、事例研究を行った。まずシステム開発分野における DOA について概念の整理を行った。次に、人文学研究における DOA についてその定義を定め、それをどのように実践するか明らかにした。また、DOA を用いることによる利点を示した。さらに、DOA による事例研究として、明治初期の物産誌である『斐太後風土記』を入力した『斐太後風土記』データベースを紹介し、GIS と統計による研究活用例を示した。

1. はじめに

データ中心アプローチ（Data Oriented Approach : DOA）とは、情報工学のシステム開発分野の用語である。データベースを中心にシステムを設計し開発を行う手法で、今日のシステム開発において広く用いられている。本稿では、この手法を借用して人文学研究に持ち込み、事例研究を行った。もちろん、むやみに他分野から概念借用を行うことは望ましくない。用語のコンテキストが異なり、誤用となることが多いためである¹。しかし、人文学に筆者の意図する事柄を指す用語がなく、また新たに造語を作ることも望ましくない。そのため、この用語を借用し、人文学研究の中に位置づけることにした。

1.1 研究の背景

近年の情報化社会の発展により、コンピュータが人文学の研究者にとっても大変身近なものになった。多くの研究者が表計算ソフトウェアなどを用いてデータを作成し、研究に活用している。研究プロジェクトでは、研究計画にデータベース

の構築を掲げ、成果の公開をうたっている。しかし、研究利用が可能なレベルで構築・公開され、継続的に利用されている人文学のデータベースは、それほど多くないようである。

問題点の一つに、人文学において、データベースに関する教育が十分に行われていないという現状がある。もちろん、専門領域が異なるため、詳しく知っている必要はない。しかしこの問題を抱えたまま構築されたデータベースは2つの道をたどる。一つは、データベースと名前のついたデータの塊が作られる。もう一つは、データベースを専門とする研究者、もしくは業者に委託される。前者の場合は、複雑化の一途を辿りそのうち管理が難しくなる。後者の場合は、お互いの意思疎通がうまく行われていれば、十分な成果が得られる。しかし、多くの場合それはうまくいかない。なぜならばデータベースの活用のためには、そのデータベースについての知識が必要だからである。

この問題を解決するためには、やはり人文学の研究者が多少なりとも、データベースについて知る必要がある。人文学の研究者がデータを扱うように変化したのだから、そのための方法論も正しく学ばなければならない。

本稿ではデータ中心アプローチという、システム開発分野の手法を借用して、人文学研究に取り

¹ 他分野からの概念借用の問題については Sokal (1998) に詳しい。

入れる事により、人文学研究におけるデータベース論を確立する。そして具体的な研究事例を示す。

2. データ中心アプローチとは

まず、システム開発²の分野において、データ中心アプローチが何を意味するのか解説する。

2.1 システム開発分野における DOA

システム開発分野における設計手法には、大きくプロセス中心アプローチ（Process Oriented Approach : POA）とデータ中心アプローチの2つがある。

ある小売店を例にこれらを説明すると、プロセス中心アプローチでは、その業務を中心にシステムを設計する。商品の販売、納入、在庫管理、売上集計など、それぞれの業務に合わせてシステムが作られる。データは、それぞれの業務から帳票として出力される。

一方、データ中心アプローチでは、データを中心にシステムを設計する。在庫や売上などのデータは、データベースに集められ集中管理される。そして業務はこのデータベースに対する IN / OUT とみなされる。すなわち、商品販売の際には、在庫が減り、売上が増えるといった具合である。

今日、データ中心アプローチにより設計された業務システムは、ありとあらゆる場面で見ることができる。スーパーやコンビニエンスストアなどのレジで用いられている POS（Point of Sale）システム、銀行の預貯金管理システム、宅配便の配達管理システム、鉄道や航空機の空席管理システム、行政、病院、インターネット、などなど私たちの生活の隅々まで入り込んでいる。

2.2 リレーショナルデータベースとは

プロセス中心アプローチとデータ中心アプローチの最大の違いは、データベースの有無である。データベースという言葉には、広義には「データを収集・整理して検索など活用しやすくしたもの」という意味がある。しかし、データ中心アプローチにおいて用いられるデータベースという用語は、この意味ではない。ここでいうデータベ

² システム開発とは一般的に、ある要件（例えば業務の省力化など）を解決するコンピュータソフトウェアを作成することを指す。

給与表

年	月	従業員	部門	給与
2011	4	山田太郎	総務部	240,000
2011	4	田中花子	総務部	250,000
2011	4	京都一郎	営業部	220,000
2011	5	山田太郎	総務部	220,000
2011	5	田中花子	総務部	230,000
2011	5	京都一郎	営業部	200,000

部門

部門コード	部門名
1	総務部
2	営業部

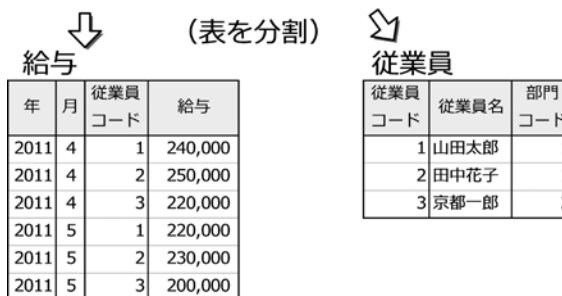


図1 リレーショナルモデルの概念例

スとは、一般的にリレーショナルデータベースを指す³。先に述べた、今日のコンピューターシステムの活躍は、このリレーショナルデータベースにより支えられている。

リレーショナルデータベースとは、リレーショナルモデルにより設計・構築されたデータベースである。リレーショナルモデルとは、1970年にIBM社のEdgar F. Coddにより提唱された⁴、集合論に基づくデータベースのモデルである。これは、大きなデータを扱う際に、冗長性を排除し、整合性を保つための方法として提案された。具体的には、データから重複した値を取り除くために、非重複の表⁵に分割し、それぞれの関係をもってデータベースとしたものである。図1に例を示す。これらの表は集合であり、原則並び順を持たない。

2.3 ERMとER図

現在、このようなリレーショナルデータベースの設計にはERM（Entity-Relationship Modeling）が用いられる。これは、1976年にMITのPeter P. Chenにより提唱された⁶データモデリングの方

³ 他に階層型データベース、オブジェクト指向データベース、XMLデータベース、キー・バリュー型データベースがある。しかし、現行の業務システムの多くはリレーショナルデータベースを用いている。

⁴ Codd (1970) で提唱された。概念は前年のCodd (1969) で示されている。

⁵ 厳密には表ではなくリレーションである。しかし、ここでは説明の便宜上イメージしやすい表と書いた。詳しくは鈴木 (1998) の第3章を参照。

⁶ Chen (1976) を参照。

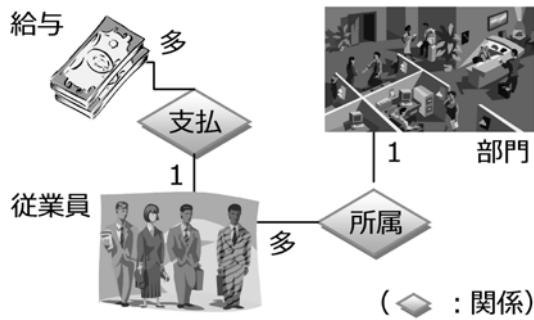


図2 ERMの概念図

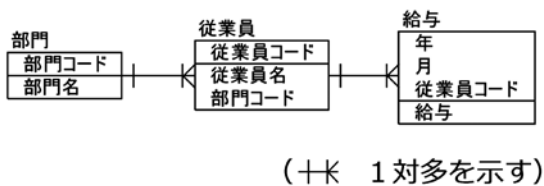


図3 Martin/IE 記法によるER図

法で、現実の世界をエンティティ（実体）とその関係で捉えようとするものである。

図2はERMの概念図である。ここには部門、従業員、給与の3つのエンティティがある。従業員は一つの部門に所属しているため、その関係は1対多である。また従業員には月々給与が支払われるため、その関係は1対多である。このように、所属、支払の2つの関係が捉えられる。

またその表現にはER図（Entity-Relationship Diagram）が用いられる。これは同じくChenにより提唱されたもので、実体を四角で、関係をひし形と線で表すものである。のちに多くの記法が提案され、主要なものにChen、IDEFIX、Bachman、Martin/IE、Min-Max/ISO、UMLなどがある。

図3ではMartin/IE記法によるER図の例を示す。エンティティは四角で表現され、図形内に列名が列挙される。列名のうち仕切りより上に、ユニークなIDである主キー（primary key）が示される。またエンティティ間の関係は線で示される。その対応関係はカーディナリティ（Cardinality、多重度）と呼ばれ、エンティティとの接続部分に表現される。Martin/IE記法では、カラスの足（Crow's Foot）と呼ばれる書き方で表現する。つまり、その関係が1対多であった場合、前者には一本線、後者にはカラスの足のような三本線で接続される。これは直感的で分かりやすいため、本稿ではこの記法を採用している。

2.4 DOAによるシステム開発

以上、データ中心アプローチとリレーショナルデータベース、またER図について概観してきた。今日のデータベースを用いた業務システムの開発では、ほとんどの場合、データ中心アプローチによるシステム設計が行われている。開発プロジェクトの最初に、顧客から業務についてヒアリングを行い、エンティティを抽出する。同時に、それらが持つべき属性と、それらエンティティ間の関係を洗い出し、ER図にまとめる。この図を元にして、実際にデータベースを構築する。そして、それにアクセスする業務を手助けするようなプログラムを開発していくのである。これがデータ中心アプローチによるシステム開発のおおよその手続きである。

データ中心アプローチ（Data Oriented Approach）という用語は和製英語である。堀内一（東京国際大学教授、当時日立製作所）が1985年に日経コンピュータの記事上で使い始めたのが最初とされている。海外ではData Centric Approachという用語が一般的である。またOOAの表記は一般的にオブジェクト指向分析（Object-Oriented Analysis）を指すため注意が必要である⁷。

3. 人文学研究におけるデータ中心アプローチ

ここでは、人文学研究におけるデータ中心アプローチとは何を指すのか定義し、またそれをどのように実践するのか明らかにする。

3.1 人文学研究におけるDOAの定義

システム開発分野のデータ中心アプローチとは、データを中心にシステムを設計する手法である。これに倣い、人文学研究におけるデータ中心アプローチとは、データを中心に研究を設計する手法を指すこととする。これはデータベースを中心としたリサーチデザインであり、各研究テーマ

⁷ オブジェクト指向プログラミング（OO Programming）とは大規模なソフトウェアを書く際にコードの不要な複雑化を避けるための手法である。システム開発において顧客の要求分析の際にOOPに配慮したのがオブジェクト指向分析（OO Analysis）、設計の際にOOPに配慮したのがオブジェクト指向設計（OO Design）である。オブジェクト指向データベースは、OOPとうまく連携できるように、データベースにオブジェクトを格納できるようにしたものである。オブジェクト指向を過度に拡大解釈してはならない。

は、データベースに対し従の関係となる。またここでいうデータベースとはリレーショナルデータベースのことを指す。

3.2 人文学研究における DOA の実践

次に、人文学研究において、どのようにデータベースを構築すれば良いかについて述べる。

データベースの構築に際し、最も重要なことは、その目的を定めることである。データベースとは、目的に応じて構築するものである。そのため、同じ対象を扱っていても、その目的によって異なる仕様のデータベースが作られる。例えば、図書館の貸出管理データベースと、eコマースなどの図書販売データベースは、同じ本という実体を扱っていながら、データベースの仕様は全く異なる⁸。

このように、データベースとはある視点から見た現実世界の一面を写しとったものであり、現実世界の複製では決してありえない。データとは、現実世界において認識された、ある対象のある部分について抽象化して、値としたものである。その観察部位は観察者に依存して無限にありうる⁹。構築の目的を定め、対象を単純化してデータベースに格納することを考えなければならない。

さて、人文学研究におけるデータベース構築の目的とは、研究活用である。ここでは、データベースの活用方法として、具体的に2つ提案したい。一つは統計分析、もう一つは地理分析である。前者については、データは質的もしくは量的データとして格納されており、大量のデータを扱うことの出来るデータベースならではの活用方法である。後者については、現実世界に存在している(していた)ものは、全て位置情報を持っている。その位置情報を用いて地理的な分析を行う。またこの2つの分析方法は、組み合わせて用いることによりさらに大きな力を発揮する。

図4に示すとおり、データ中心アプローチにおけるデータベースの研究活用とは、具体的に統計分析と地理分析を指すこととする。もちろんこれ

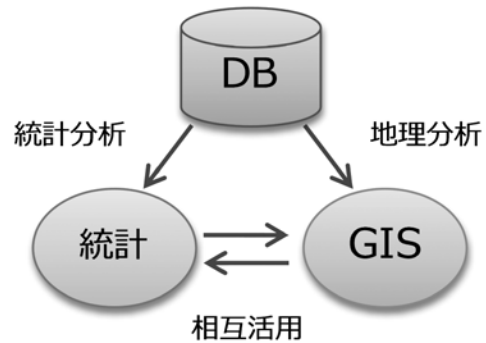


図4 DOAにおけるデータベースの活用

は提案であり、この他にも活用の方法があれば加えてもらって構わない。

他にデータベース構築の際に重要な点は、可能な限り小さく作ることである。往々にしてデータベースは肥大化する。欲張って属性を沢山取ってしまうのである。そして、そのデータを入力しない。時間が限られていたり、仕様に無理があったり、また重要で無いために未入力もしくは正しい値が入力されない。そのようなデータが1件でも含まれると、データベースの価値は大きく損なわれる。誤りを含むデータベースは、第三者が信用して使うことができないためである。

データベースを小さく作るためには、その利用目的を明確にしておくことが重要である。そして、その目的を達成する、最小サイズのデータベースを構築する。つまり、目的を達成するだけの、最少のテーブル¹⁰と属性(例)、そしてその主張に十分なデータ(行)の件数があれば良いのである。

そして、小さなデータベースを手早く作り、研究活用することが大切である。データベースのテーブルや属性に不足があると気がつけば、仕様変更を行う。実は、データベースは正しく設計してあれば、その仕様を変更するのは難しくない¹¹。入力のコストに比べ、はるかに容易で

⁸ 例えば、前者は本の価格や重量についてのデータを保持しない。売値や送料について考慮する必要が無いためである。

⁹ また、観察の理論負荷性 (theory-laden) の問題もある。これは、Norwood Russell Hansonにより提唱され、観察とは観察者の持つ知識や視点に依存しており、完全に客観的な観察はありえない、とするものである。Hanson(1958)を参照。アヒルとウサギの多義図形(ゲシュタルトスイッチ)がよく知られている。

¹⁰ テーブルとは、データベースで用いられる2次元の表を指す。

¹¹ 一般的に、リレーショナルデータベースの仕様変更は困難である、とされているが、それはシステム開発においての話である。データベースを複数のソフトウェアから呼び出しているため、影響が大きいという主張で、これは全く正しい。しかし、人文学で研究利用するデータベースは、利用者がそれほど多く無く、また専用のソフトウェアも作らない事が多いため、影響は少ない。可能な限り小さく作り、活用して、仕様変更を繰り返すのがよい。またこの点より、XMLデータベースに優位性があるという議論は、ここでは成り立たない。

ある。いつまでも仕様を考えたり、使うか分からない属性を入れ続けるより、最小サイズで作って雪ダルマ式に大きくしていくのが、良いデータベースを作る要点である¹²。

3.3 データ中心アプローチによる利点

データ中心アプローチから得られるメリットは3つある。一つ目は、データを管理しやすい、という点である。リレーショナルデータベースには重複するデータが無く、また複数人での同時アクセスに対応しているため、データの一元管理が可能である。またバックアップのための仕組みが備わっており、データの永続化を担保できる。

二つ目は共有可能、という点である。従来それぞれの研究テーマに対しデータベースが個別に作られてきた。データ中心アプローチではデータベースを主に、各研究テーマ、研究者を従の関係とすることにより、データベースを共有することができる(図5)。同じデータベースの利用、もしくはデータベース間の比較を通じて、研究者間の研究交流が促進される。またそれぞれの研究テーマにデータベースを通じて繋がりが形成される。その際、それぞれのデータベースは無理に融合させる必要はない。個別の研究に最適な形で保持し、それら複数のデータベースに共通する項目をもって緩やかにリンクさせれば良い¹³。

三つ目は公開しやすい、という点である。今日インターネット上で利用されているデータベースのほとんどは、リレーショナルデータベースである。そのため、研究で構築したデータベースを

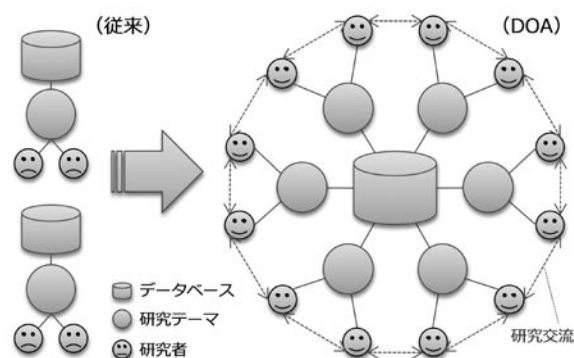


図5 DOAによるデータベースの共有

そのままインターネット上にアップロードすることができる¹⁴。公開の際にはデータ表示のためのサイトを作る必要があるが、それは難しくない。ホームページ作成の一環として専門の業者や協力者に頼むこともできる。データベースの公開は研究成果の社会還元として意義が大きい。

研究活用を行ったデータベースは積極的に公開すべきである。第三者が、公開されたデータベースを活用する際は、必ず先行研究を参照する。構築の経緯や信頼度を確認するためである。今日、論文の学術的価値は被引用回数(citation)により測られる。データベースを公開することは、その研究の学術的評価を高めることに繋がるのである。データベースをそのまま手元に置いておくことは、ただ死蔵を招くだけである。

4. 事例研究

ここでは人文学研究におけるデータ中心アプローチの実践例として、『斐太後風土記』データベースを用いた事例研究を紹介する。

『斐太後風土記』とは明治六年(1873年)に完成した岐阜県飛騨地方の地誌である。編者は飛騨の地役人である富田禮彦で、飛騨地方の415村落について人口、戸数、産物等について記載している。この資料からは、産業化以前の日本の村落の実態について知ることができる。本資料の重要性については以前より知られており、国立民族学博

¹² これは、システム開発における反復型開発(Iterative and incremental development)のアナロジーである。ウォーターフォール型の開発モデルに對置され、アジャイルソフトウェア開発、エクストリーム・プログラミング(XP)が知られている。筆者はこれらの信奉者という訳ではないが、人文学研究におけるデータ中心アプローチでは、反復(イテレーション)によるリスク最小化は取り入れるべきと考えている。

¹³ それぞれのデータベースは個別の研究に最適な形、つまり個別の研究を行う上で最小のサイズで保持する。多くのデータベースでは、データベース、スキーマ、データベースオブジェクト(テーブルなど)という三層構造を持っている。一つのデータベースに複数のスキーマを、一つのスキーマに複数のテーブルを持つ構造になっている。ここでいう個別研究に最適な大きさのデータベース、とは、スキーマに対応する。研究目的ごとにスキーマを作り、それらを緩やかにリンクさせるのがよい。また、データベースのリンクのための共通項目とは、抽象度の高いものを用いる。例として、水田耕作、焼畑、漁業、などの人間活動が挙げられる。また位置情報も共通項目となる。

¹⁴ 今日では年間2,500円ほどでホームページスペースとデータベースを借りることができる。スペース上ではPHPやPerlなどが動作し、データベースはMySQLやPostgreSQLを使うことができる。維持費や停電などの障害対策、データのバックアップを考えると、レンタルするのも良い方法である。

物館では1978年以降コンピュータを用いた計量的研究が進められ、成果がまとめられている¹⁵。

4.1 『斐太後風土記』データベース

データベースの構成については図6の通りである。郡、郷、村落、産物、品目、品目親、品目タイプの7テーブルを作成し、非重複のデータを格納した。村落テーブルには『斐太後風土記』に記載されている415村落が格納されている。郡、郷テーブルには、飛騨地方の郡および郷の一覧が格納されている。品目テーブルには『斐太後風土記』に記載されている産物の品目名を、非重複で格納した。産物記載には、一つの産物についてアユ、鮎、年魚のように表記ゆれを持つものがある。この表記ゆれを吸収し、親名をつけたものを品目親テーブルに格納した。さらにこの品目親について穀類、種実類、魚類などタイプ分けしたものを品目タイプテーブルに格納した。また村落テーブルと品目テーブルをリンクさせ、村ごとの各産物の数量と単位を産物テーブルに格納した。

データベースのソフトウェアにはOracle Database 10g Express Editionを採用した¹⁶。またGISでの地理分析を行うために、415全ての村落について「陸地測量部五万分一地形図」を利用して村落位置を特定し、緯度経度の入力を行った¹⁷。

4.2 研究の背景

データベースを用いた『斐太後風土記』の研究には、国立民族学博物館の小山修三らが1978年以降に行った食料資源の計量的研究がある¹⁸。この研究では、山村における食料生産という視点から産物の食料品目に着目し、村あたりの食料品目の総エネルギーを人口で除して栄養学的考察が行われた。しかし、養蚕や製糸などの手工業生産から得られた収入については検討が行われなかった。当時飛騨では、ほぼ全ての村落において養蚕が

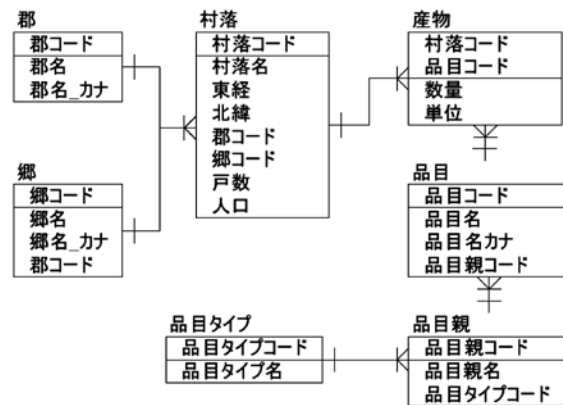


図6 『斐太後風土記』データベースのER図

行われていた¹⁹。繭から作る生糸は、当時の飛騨の主要な輸出品であり、総輸出金額の実に79%を占めていた²⁰。また輸入品の30%は米、9%は雑穀であり、大量の食糧買い入れを行っている²¹。

『斐太後風土記』の大きな特徴は、産物の品目名だけでなく、数量も記載している点である。小山らの研究では、産物収量をエネルギー(kcal)に換算することにより、異なる種類の食品を横断的に比較可能にした。本研究では、産物の収量を金額に換算することにより、食料品目に加え手工業製品についても比較可能とする。そして人々が何れの産業に力を注いでいたのか、明らかにする。対象は図7に示す庄川流域の42村落とする。

庄川流域の村落を表1に示す。庄川流域の上流(南部)の19村を上白川、下流(北部)の23村を下白川という。福島村の南に「福島歩危」と呼ばれる交通の難所があり²²、ここがおおよその境となっている(図8)。庄川流域の42村落の産物のうち、食品は74品目、手工業製品は31品目である。ここでは、このうち主要な産物である米、稗、繭の3品目を取り上げる。稗は42全ての村

¹⁵ 秋道(1979)、小山ほか(1981)、小山(1984)、藤野(1982)、松山(1979)、Koyama, S; Thomas, D. H. (1981)など。

¹⁶ 一部性能に制限があるが、無償で利用できるDBMSである。4GBまでのユーザー・データ領域、単一のデータベース・インスタンス、単一CPU、1GBまでのRAMの制限がある。

¹⁷ 明治42年測量の西赤尾、43年測量の白川村、白山、白鳥の図郭を用いた。国土院の提供する「ウオッチズ」<http://watchizu.gsi.go.jp/>を用いて、現在の地形図と比較しながら位置を入力した。

¹⁸ 小山ほか(1981)を参照。

¹⁹ 403村(97%)において繭の産物記載がある。

²⁰ 『斐太後風土記』の「國産諸品賣出價概記」(明治三年)による。絲6,750貫目、255,000両である。

²¹ 『斐太後風土記』の「必需品他國より買入高凡積」(明治三年)による。米15,000石、88,230両餘である。また雑穀8,000石、26,660両餘である。

²² 『斐太後風土記』の福島村に次の記載がある。「福島歩危福島村と尾神村との間なる、岩山の絶壁を研割て路を作れり。郷中にも、国内にも、比類なき險難の歩危路にて、鬚摺・鞆丸縮等の名に負ふ難所あり。雪中は皆長瀬の枝村秋村へ渡り、其の崖路を避けて行通へど、是亦一本九繼の長橋ありて、中間にて自然動搖ぬれば、其危きことは、薄氷を踏がごとし。」

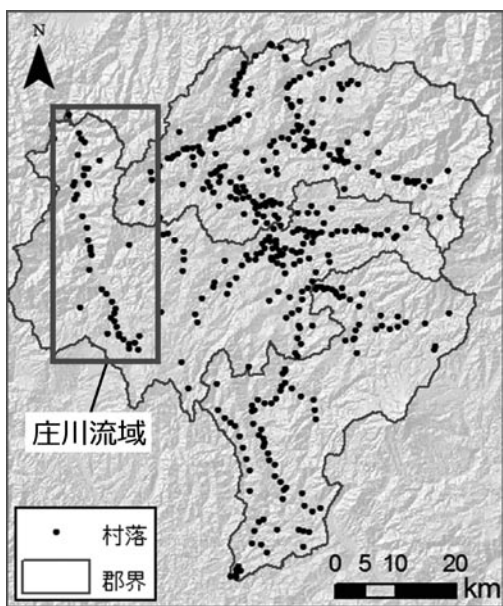


図7 『斐太後風土記』記載の村落

で生産され、収量も4809.5石と随一の産物である。繭も42全ての村で生産され、収量は5275.2貫目と極めて高い。米は32村(76%)と全ての村では生産しておらず、また収量も752.85石と高くない。しかし、飯島村では119.8石を生産するなど、偏りが大きい。また先に触れたように、輸入品目の30%を占めるなど、明治初期においても米食への志向は強かったようだ²³。

4.3 米・稗・繭の主成分分析

これら主要な生産物である米、稗、繭についてその生産量に着目し、村落がいずれの産業に力を入れているか分析した。具体的にはそれぞれの生産量を一人あたりの金額(両)に換算し²⁴、主成分分析を行った²⁵。

結果を図9に示す。第一主成分は養蚕または稗

²³ 『斐太後風土記』の中畑村に次の記載がある。「然るを後世になるまゝに、村民さがしらに成て、稗のみ作らむよりはと思ひ、何れの村も辛して河水を堰上げ、水田になして、稻をも聊か作れる事になりしは、や、後の事にならむ。元來寒冷なる地理なれば、稗のみ作れば、凶年にも、少しは秋成もあるべきを、強て稻を作故に、早霜の年にはみのらず、自然、飢に及ぶこと多し。可憐。」

²⁴ 金額換算は米、稗については『斐太後風土記』の「必需品他國より買入高凡積」より算出した。米は5.882両/石である。稗は雑穀の値を用い、3.3325両/石である。繭については「蝸萬日記帳」より明治五年の値を用いた(『朝日村史』第3巻 p.441)。1.2621両/貫である。また人口の影響を取り除くため、それぞれの金額を人口で除した。

²⁵ 分析には統計解析ソフトウェアRを用いた。http://www.r-project.org/ パッケージはFactoMineRを用い、主成分分析には関数PCAを用いた。

表1 庄川流域村落の産物と人口

村名	米	稗	繭	人	村名	米	稗	繭	人
六蔵	0	284.1	30	90	尾神	2.8	28.5	49.1	60
三尾河	0	145	32	80	福島	0.1	5.5	44.6	10
寺河戸	1.2	400.2	98	230	牧	1.6	16	36.3	30
黒谷	4.5	360	58	160	長瀬	8.9	128	218	220
総則	10.5	155	48	90	御母衣	2.8	22	187	70
一色	29.8	330	75	120	平瀬	1.45	62	143	110
猿丸	10.5	87.1	35	80	木谷	3.1	57.2	184.8	140
町屋	65.2	200.2	64	90	俣木脇	1.6	16	18	60
新淵	79.2	220	89	160	野谷	3.6	10.2	22	20
野の俣	9.5	276	73	180	大牧	34	28	9.3	100
中畑	32.5	134	58	100	萩町	105	2.3	970	390
牧戸	19.8	77.6	48	40	島	15	33	300	70
牛丸	35.2	114	68	100	牛首	0	30.2	85	40
岩瀬	30.5	160.5	60	90	鳩ヶ谷	50.8	42.5	180	100
赤谷	25.6	390	62	180	飯島	119.8	150	600	350
森茂	4	65	25	30	大窪	0	2.4	16	20
中野	13.5	220.5	75	130	馬狩	0	40	22.5	60
海上	15.5	240	110	120	内ヶ戸	0	33.4	150	30
尾上郷	0	48.3	30	30	加須良	0	45.3	180	60
(六蔵～尾上郷を上白川、尾神～小白川を下白川という。『斐太後風土記』の記載による)					椿原	0	41.5	220	40
					有家ヶ原	5.3	25	191.8	50
					芦倉	0	43	209.8	60
					小白川	10	40	100	60

(単位：米、稗は石、繭は貫)



図8 庄川流域の村落と福島歩危

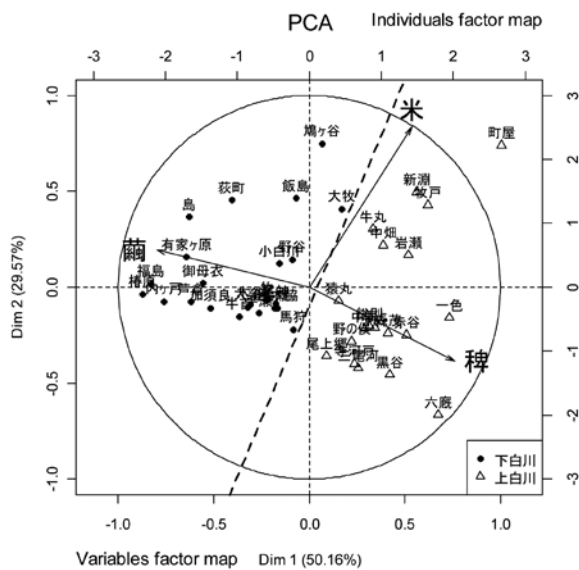


図9 米・稗・繭の主成分分析

生産への依存度を表している。第二主成分は米作への依存度を表している。また村落は、第一主成分軸方向に、上白川、下白川の群に分かれた。図中の破線は上・下白川区分の判別分析による。

産物からは、下白川の村落は繭の生産に、上白川の村落は稗の生産に力を注いでいることが分かった。繭も稗もこの地域の42全ての村落で生産されている。しかし、生産金額は下白川が繭に、上白川は稗に偏っている。

また南北関係なく、一部の村落は米の生産に力を入れていることが分かった。これは平坦である、標高が高すぎない等、水田に適した土地を多く持つ村落と推測される。

4.4 考察

この分析結果について考察を行う。当時、山がちな飛騨の物資の運搬は「歩荷」と呼ばれる人の背によるものであった。重量物である米の運搬は困難を伴ったと考えられる。しかし一方で、当時の飛騨の人々は米食への強い志向があった。そのため、下白川の村落で繭を売ってお金を手にした人々は、さらに北の、米どころである富山県より運搬された米を購入して食べたと考えられる。しかし、東西南方を山で囲まれていた上白川の村落では、お金があっても重量物である米が運搬されず、そのために稗を中心に生産する暮らしが残ったと考えられる。つまり、「下白川の村が繭に、上白川の村が稗に偏ったのは、富山県との位置関

係が原因」という仮説を提示することができる²⁶。

5. おわりに

以上、データ中心アプローチの人文学研究への位置づけと事例研究について示した。再度述べると、人文学研究におけるデータ中心アプローチとは、データベースを中心に研究を設計する手法を指す。そして、その活用には統計分析と地理分析が用いられる。その際のデータベースとは、リレーショナルデータベースを指す事とする。

従来、データベースという言葉の多義性に問題があった。広義の意味では何でもデータベースになってしまう。これが日本の人文学研究において定量的研究の発展を阻んでいる最大の原因と推測している²⁷。人文学研究におけるデータ中心アプローチの提唱により、このような現状に一石を投じることが出来れば、望外の喜びである。

また、本手法は人文学研究であれば分野を問わず適用可能である²⁸。研究事例としては筆者が行っている物産誌の例のほか、中村大の進めている考古学研究における例がある²⁹。

データベースについては、多くの良書がある。分かりやすく厚くない入門書として鈴木(1998)を挙げておく。

謝辞

本稿の執筆にあたり、同志社大学文化情報学部の矢野環教授からは統計学およびR、また文献の読み解きについて、多く指導を受けた。深謝の意を表す。また、総合地球環境学研究所の中村大研究員とは、本稿執筆の元となる議論を重ね、多くの教示を受けた。深謝の意を表す。

²⁶ この検証はデータからは行えないが、文献史学など他分野と連携して考えていくことができる。矢ヶ崎(1958)の近世末期における口留番所の研究によれば、庄川流域北端の小白川口では、米が34.36石/月、稗が0石/月の移入がある。これに対し、南端の野々俣口では、米が0石/月、稗が7.9石/月である。同じく南端の寺河戸口では米が0石/月、稗が5石/月の移入である。

²⁷ データベースが正しく作れないため、データを用いた定量的研究が継続的に行われない。また公開が進まないため、共同利用が行われない。

²⁸ 自然科学、社会科学においては、人文学とは異なり、サンプルを多く取れるという違いがある。それぞれの領域に適した方法が必要である。

²⁹ 松森(2010)、中村(2011)を参照。

本研究は総合地球環境学研究所「東アジア内海の新石器化と現代化：景観の形成史」プロジェクトの成果の一部を利用している。

参考文献

- Chen, P. P. (1976). The Entity-Relationship Model -Toward a Unified View of Data. *ACM Transactions on Database Systems* 1 (1): 9-36.
- Codd, E.F. (1969). Derivability, Redundancy, and Consistency of Relations Stored in Large Data Banks. *IBM Research Report*, San Jose, California RJ599.
- Codd, E.F. (1970). A Relational Model of Data for Large Shared Data Banks. *Communications of the ACM* 13 (6): 377-387.
- Hanson, N. R. (1958). Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science. Cambridge University Press.
- Koyama, S; Thomas, D. H. (1981). Affluent Foragers. *Senri Ethnological Studies* 9.
- Sokal, Alan; Jean Bricmont (1998). Fashionable Nonsense. New York: Picador.
- Sokal, Alan; Jean Bricmont 著 (1998). 田崎晴明, 大野克嗣, 堀茂樹 訳. (2000). 『「知」の欺瞞 — ポストモダン思想における科学の濫用』岩波書店
- 秋道智彌 (1979) 「明治初期・飛騨地方における生産魚類の分布論的研究」『国立民族学博物館研究報告』4 卷 1 号
- 朝日村誌編集委員会 (2005) 『朝日村史』第 3 卷朝日村
- 小山修三・松山利夫・秋道智彌・藤野淑子・杉田繁治 (1981) 「『斐太後風土記』による食糧資源の計量的研究」『国立民族学博物館研究報告』6 卷 3 号
- 小山修三 (1984) 『縄文時代—コンピューター考古学による復元』中央公論社
- 鈴木健司 (1998) 『データベースがわかる本』オーム社
- 中村大 (2011) 「祭祀考古学における多変量解析と GIS の活用」『國學院大學伝統文化リサーチセンター研究紀要』第 3 号
- 藤野淑子 (1982) 「明治初期における山村の食事と栄養—『斐太後風土記』の分析を通じて—」『国立民族学博物館研究報告』7 卷 3 号
- 松森智彦 (2010) 「RDB による『斐太後風土記』データベースの構築と活用」『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』
- 松山利夫 (1979) 「明治初期の飛騨地方における堅果類の採集と農耕」『国立民族学博物館研究報告』4 卷 1 号
- 矢ヶ崎孝雄 (1958) 「飛騨における商品流通の地域的

研究—近世末期の場合—」『金沢大学教育学部紀要』第 6 号

附 録

本稿の主成分分析の作図 (図 9) に用いた R のコマンドを記しておく。data.csv は、表 1 を元に本稿中の金額換算比を用いて作成した、村名, 米, 稗, 蕪の 4 列 43 行の金額データとする。

```
library("FactoMineR")
library("MASS")
data<-read.csv("data.csv", header=T, row.names=1)
symbol<-c(rep(2, 19), rep(16, 23))
res<-PCA(data, graph=F)
plot.PCA(res, choix="var", title="", cex=1.8)
par(new=T)
plot(res$ind$coord[,1:2], xlim=c(-3,3), ylim=c(-3,3),
      pch=symbol, xlab="", ylab="", axes=F)
text(res$ind$coord[,1:2], labels=rownames(data),
      pos=3, cex=0.8)
axis(3)
axis(4)
legend("bottomright", legend=c(" 下白川 ", " 上白川 "),
      pch=c(16, 2))
mtext("Variables factor map", side=1, adj=0,
      line=3, cex=1.2)
mtext("Individuals factor map", side=3, adj=1,
      line=3, cex=1.2)
mtext("PCA", line=2.7, cex=1.5)
lda<-lda(symbol~ res$ind$coord[,1]
          + res$ind$coord[,2])
const<-mean(lda$means %*% lda$scaling)
a<-const / lda$scaling[2]
b<- -lda$scaling[1] / lda$scaling[2]
abline(a, b, lty=2, lwd=2)
```