

データベース活用におけるバザール型コミュニティ形成について

About the Bazaar Type Community on Database Utilization

藤岡 洋

Hirosi Huzioka

東京大学 東洋文化研究所, 東京都文京区本郷 7-3-1

The University of Tokyo, 3-7-1 Hongo Bunkyo, Tokyo

あらまし: 人文系研究機関でデータベースの乱立が始まってから早 15 年近くたつ。今回はグローバル COE、21 世紀 COE の前身、COE 形成プログラムで構築したデータベースを振り返る。その上で、なぜデータベース継承が難しかったのか、問題を整理したい。さらに、データベース活用は幅広い利用という観点とそれと関連する技術が主流を占めているように見受けられるが、今回はバザール型コミュニティを仮定し、その視座から上記2点の可逆的に考えてみたい。

Summary: This paper attempts to explain why database succession has been difficult. At the moment, almost 15 years after the initial increase and flooding of databases have passed. In many ways there is a strong case for stopping their activity, i.e. because of the lack of concept for a programmer or a manager in manual documents or technical reports. In this paper, I elaborate the meaning of a "player" in database utilization. In order to do so, I present an overview of "concept map for database" and "bazaar type community on database utilizations".

キーワード: 人文科学, データベース

Keywords: humanities, database

1. 大規模科研でのデジタルアーカイブ構築を振り返る

平成 11 年度から 15 年度までの 5 年間、東京大学人文社会系大学院には象形文化研究拠点(以下、拠点)が設置され、その主要プロジェクトの一つとして国内外の古代ローマ史考古学者、美術史家、情報学者などが協力してデジタルアーカイブを作る試みがなされた。このプロジェクトは文部科学省 COE 拠点形成プログラムとして実行された。COE 拠点形成プログラムは、グローバル COE、その前身である 21 世紀 COE のさらに前身となるプロジェクトであった。この拠点形成プログラムの目的は次のようなものであった。

「高度情報化社会においては、情報科学の技術的進展に応じて情報通信基盤の整備だけでなく、情報内容(コンテンツ)が重視されるようになるのは必然であるが、近年喧伝されている「コンテンツ重視」の主張は、ともしれば、情報の量的増大のみを招来するに留まっている。他方、情報の出典の検討と情報の質の検証という観点からみれば、人文学は対象とする原資料ごとにその基本的操作法を方法論として確立している。本研究は以上のような現状認識に立って、高度情報

化社会において最も重要なものとなる情報のコンテンツに対して、基礎研究である人文学の立場から寄与しようとするものである。研究の機軸とするのは、人文学がその基盤として確立してきた原資料批判の方法論(史料学・文献学を含む資料学)を再認識して統合し、その総合の上に立ってコンテンツの学としての情報学と融合させることにある。そのために、これまでの人文学で確立されてきた原資料批判の方法論とその成果の蓄積である学術情報を、人文学の基本的な方法論の観点から更なる体系化と統合を行い、象形文化アーカイブへと集約する」[1]

1.1 アーカイブの概要

ポンペイデジタルアーカイブはこの拠点プロジェクトの中心的なコンテンツであった。20000 枚ほどのポジフィルムから大型スキャナ(Heidelberg, Imacon など)を使ってのデジタル化と PPM(Pompei Pitture e Mosaici)とそれを元とした最新の文献研究成果の統合を目指した。データベースの果たす役割は、いうまでもなく、この「統合」部分であった。筆者は創設当初には拠点に所属していなかったため詳しい経緯は知らないが、日

立制作所の DIS 開発と同時並行的になされており、すでにプロトタイプが出来上がっていた。[2]



図 1. 日立 TSS

図 1. のとおり、当初のアーカイブはスタンドアロンのタイプであった。おそらく博物館などでの展示公開を念頭に置いていたと思われる。その仕様は、2000 年初頭としては珍しくタッチパネルを採用しており、利用者はローマ帝国全図から地図階層をたどっていき、イタリア、ナポリ、ポンペイ、地区、建物、部屋、壁、最終的にはこの領域では標準となっている 9 分割区分の領域画像一覧が画面下 2 分の 1 に表示される。そのサムネイルをタッチすることで地図から拡大画像に切り替わり、サムネイル表示部分に PPM 情報やそれに基づく最新の学術情報などの各種学術情報がタブ形式で表示される、というものであった(参照: 図 2)



図 2. 拠点アーカイブ(例: 最下層)

ただこのスタンドアロンタイプは仕様上というよりも、構築プロセスに大きな問題があった。画像の著作権について日立に全面的に譲渡することに躊躇した拠点側が、それをコントロールするために、1) 候補を挙げ、2) 合議にかけ、3) CD-R に焼いて日立に送るという方

法が採られていた。日立側でデータベースに載せる作業に数週間から数ヶ月を有した。

ここからアーカイブ構築を完全に拠点側で構築する決定がなされたのは拠点形成後 3 年以上たってからのことであった。

1.2 システム

拠点の選択肢は日立のデザインを再現してウェブブラウザ上に載せるものであった。

日立との連携で生じた問題を解消する目的もあって、すでにサーバの管理運用の補助を委託していた外部業者に文字検索インターフェイスを委託し、データ(画像とテキスト)入力インターフェイスを別の関連業者に委託した。委託業者との関係は、日立と違い、サーバ管理者の知己であり、問題だった構築に関する障害はなくなった。アーカイブ本体は古代美術史の専門知識を持つ拠点スタッフによって一人で構築された。それぞれ PSS (Pompei text Searching System), DIS (pompei Data Input System), TopV (pompei TOPographical Viewer) として新たな船出となった。アーカイブに載せられるデータ内容もこれに伴い急速に充実していった。年数回、イタリアから研究者を招聘し、集中的にテキストデータ、画像データそのものについて深夜まで議論がなされ、その結果は以前とは比較にならないスピードで更新されていった。

だが、スピードがあがると同時に、次の問題が生じた。すでにこの頃は最終年度に入ろうとしていて時限との戦いがすでに始まっており、一言で言えば「もっとスピードを」ということであった。

またウェブベースに載せたことで公開に際しての準備も潜在的な問題として残されていた。以前日立との間で生じ、その後保留したままになっていた著作権の問題だった。ポンペイには江戸時代と同じほどの長さの発掘の歴史があり、かつ未だ発掘作業が行われている。ゆえに、拠点のアーカイブには非公開状態の画像も多く載せられており、これをそのまま公開してしまっているのかについては議論が深められる余裕がなかった。

これらの諸事情を抱えたまま拠点形成プログラムは最終年度を迎えた。ポンペイデジタルアーカイブはいわゆる「落ち穂拾い」を必要とする状態で最後の国際シンポジウムを主催した。幸いなことに、その後研究代表者であった青柳正規先生(現:文化庁長官)が国立西洋美術館館長として着任されるという契機と、東京大学学環情報学府の坂村健先生のご好意もあり、東京大学病院分院跡地の研究室でアーカイブ構築は継続されることとなった。

この時期には時限との戦いに加えコストの問題が生じている。前任サーバ管理者の知己によって作られた上記2つのシステム (PSS, DIS)の改良、メンテナンス・機能追加要望に割くコストはこの時期の拠点にはすでに難しい問題になってしまっていた。PSS については、php+postgresql という当時のスタッフで十分メンテナンス・改良に耐えうる環境で書かれていたため、問題は最小限に抑えられたが、DIS については、Microsoft Visual Basic (以下、VB) で書かれていた。当時、VB に明るいスタッフがいなかったため、データ入力・編集に関する要望に応えられなくなってきた。このため、DIS を再度書き換えることにした。この部分は筆者が担当し、PSS, TopV と同じ php+postgresql で新たに書きなおした。DIS は DECS (pompei Data Edit ComponentS)として生まれ変わった。この間にすでに2年が経過している。

DECS は単に研究者のためのデータ入力システムを目ざしてはいなかった。TopV ではコンテンツの内容・配置について自由に変更ができなかったし、これまでの経緯を踏まえると人材に頼るシステムはいずれ破綻する恐れからだった。そこで、DECS は TopV すなわちコンテンツそのものにアクセスをして、すぐさま研究者の研究成果をコンテンツに反映できるようにしていった。

だが、拠点はその後1年をもって終了する。COE 拠点形成プロジェクト開始から数えれば計8年にも及ぶプロジェクトだったが、内実は試行錯誤の連続であった。しかもすでに同時期、イタリアでは同様に画像を全面に押し出したデータベースが幾つか公開されており、その数年後には Google map / Google earth によってポンペイを利用してそれをパノラマ化して散策できるようにさえなっていた。拠点のコンテンツはこれらの事例と理念からしてもまったく異なるもののはずだったが、それを差別化する方策を拠点スタッフでは当時見いだせなかったし、また見いだしても実行に移すことはもはや不可能であった。

拠点消滅の直接の原因は時限とコストという外的なものだったが、拠点アーカイブ独特の事情として画像の使用許諾の問題があった。この手の問題は現在のデータベースにはあまり見られないので今後のデータベース開発には参考にならないであろう。その一方で、データベース継承については、いくつか反省すべき点が残されたままだった。一つは継承そのものに対する考え方である。拠点アーカイブは学術的価値、データの規模からして当然学術世界から注目されるであろうという甘い見込みがあったことは否めない。しかし、これはデータベースを構築するすべてのプロジェクトに

当てはまるのではないだろうか。その先には継続利用がイコール、コンテンツの存続につながるという甘い見込みにつながっている。確かに、継続して利用されるウェブサービスはその技術はもとより、利用され続けることがその存続の大事な要素になってはいるだろう。しかし、学術データベースにとってそのような形での存続は必ずしも求められていないし、開発者も求めている場合が多い。もう一つはデータベースの完成度といったときの意味である。これは本稿の意図の中核部分であるが、データベースの完成度は基本利用者本位で図られる。だが、それがプログラム言語で書かれているかぎり、データベースがその後置かれる状況如何にかかわらず、再生可能なデータベースこそが継続性を備えたものなのであり、そうでないものはデータベース開発者が考える以上に深刻な問題をすでに抱えてしまっていると言わざるを得ない。そして我々拠点スタッフにはそのような発想がそもそもなかったのである。

2. 現在のデータベース活用に関する所見

とはいえ、いま述べてきた状況は2013年現在でも散見できる。相変わらず科研費データベースでも、データベースの構築、という文字が踊っている。これらのデータベースが数年後、十数年後停止に追い込まれるかもしれないという予測に対する対策に関するアイデアを目にすることは少なく、現況に対して楽観的すぎはしないかと心配になる。

ただデータベースの保存ないし継続的運用を視野に入れたデータベースの活用についてのアイデアは、少しずつ裾野を広げつつあるようにも見受けられる。

- 公開方法のバリエーション。モバイル環境の変化に応じたデータベースの利用範囲の拡張。
- 出版。データベースとの紐付けに関して。
- ワークショップ。公開に留めない積極的な活用方法の模索。
- データベース同士の連携など。

前者2つに関しては拠点アーカイブでも試みていたことだったので、紹介しよう。

2.1 公開方法のバリエーション

拠点アーカイブでは、最終年度にある意図をもって、二つの実験を行っていた。その意図とは、アーカイブを外に持ち出すことで、研究のフィールドにより近いところでのデータベース活用の可能性を探ることであった。

そのために、まず 2009 年に国立西洋美術館で開催された「古代ローマ帝国の遺産」展 [3] でのギャラリートークを利用して予備的な実験を行った。



図 3. 国立西洋美術館でのコンテンツ配信実験

図 3.には二つの PC が見られるが、実は演壇の講演者が説明に使っている TopV は、手前で聴講している女性の PC から発信されている。

この実験が成功して後、続いてポンペイに出向き同様の実験を行った。この時は、学術データベースが外に持ち出されることで得られる効果を考察するためだった。ポンペイでは研究者が指導のため学生を連れて公開停止中の遺構を訪れるが、その際の補助ツールとして、TopV を野外に持ちだしてもらったわけである。実際にこの時点でデータにして検討すべき課題が幾つか見つかかり、その場で大学院生と教員とで議論がなされるといった光景が見られた。これをメモする学生がおり、編集に関して柔軟性を増していた DECS を利用すれば、データ監修そのものに役立つであろうという新たな展望も寄せられた。この実験は成功だったというよいが、残念なことにその後すぐ拠点アーカイブは停止してしまった。

一つ脱線するならば、拠点が閉鎖されてまもなく、2010 年に初代 iPad が発売されたが、その宣伝で、ポンペイの遺構(とキャプションにあった)に入って土まみれの作業員が iPad を見て何かを確認するという画像(参照:図4)が使われていた。

これを見た瞬間、まさに著者が想像していた図が目の前に提示されたことに衝撃を受けるとともに、明らかにこれは問題のある嘘の広告だと分かった。少なくとも同時ポンペイに公衆無線 LAN は届いていなかったからである。しかし、いずれにせよ、同様の使い方は当時の iPad 開発者たちも考えていたという、何か背中を押されたような感傷とも、拠点閉鎖の寂寥感ともいえない複雑な気持ちになったことを覚えている。



図 4. 初代 iPad 発売時の広告

2.2 データベースとの紐付けされた出版

学術データベースの活用法として、次に挙げられるケースが多いのはデータベースの資産を使っの論文執筆、著作刊行であろう。



図5. POMPEI INSULA OCCIDENTALIS P.167

拠点でも最終的に両国の研究者が合同で取り組み一つ大きな成果をあげている。Masanori Aoyagi -Umberto Pappalardo "POMPEI (REGIONES VI-VII) INSULA OCCIDENTALIS -Volume Primo-", 560 pages, Valtrend Editore, 2006 である。

この著作には一つ、特徴があった。使用された画像(モノクロ・カラー)には、拠点アーカイブで監修・管理していた 16 桁の固有 ID が振られている。この刊行物から他の研究者が、オリジナルの画像情報に接しようとすれば、拠点アーカイブにアクセスし、その他関連

情報は誰でも調べることができた(参照:図5)。このことはアーカイブと出版物はセットとして扱うことができる可能性を示そうとしていたわけである。データベースが継続運用されていれば、さらに出版物そのものが学術成果の記録(ログ)として位置づけられたであろう。

しかし、プロトタイプとしての著作物刊行は単発で終わってしまった(そのタイトルどおり、予定では第2巻、第3巻が少なくとも予定されていた)こと、そして繰り返すが上記著作物もデータベースが運用を停止してしまったことで狙った効果が半減してしまった。現在ではこのIDの意味は消失しているわけである。データベースが停止すると書物の意義までが落ちてしまう現代ならではの現象の一つであろう。

2.3 アウトリーチとしてのワークショップ

それでもデータベース活性化の手段としてそれをつかった論文・著作活動を後押しすることが有効であるとするならば、もっとドラスティックに考えることもできる。すなわち、データベース使用を登録制にして、必ず論文ないしレポートの形でフィードバックを要求する、というものである。だが、まず利用者はタイトルだけでどのようなデータが収められているかも分からないようなデータベースに登録などしないだろうし、先に中身も見せないうちからフィードバックを求めるようなハードルに真っ向から付き合ってくれとは到底思えない。

ところが、学術データベースを質で測るという観点に立てば、次のような試みは魅力的である。著者が現在所属している東京大学東洋文化研究所には10年近くの運用実績のある「アジアンパロメータ」というアジア比較社会学のためのデータベースがある(参照:図6)。このデータベースはアジア各国に住む20歳以上69歳以下の全ての成人を母集団として対象とし、基本的に調査員による面接聴取法によって有効回収数が必要サンプル数程度になるように、各国ごとに期待される回収率を勘案して標本規模を800から2,000サンプル採り、層化多段無作為抽出法と割当抽出法を用いて、1.社会基盤整備、2.経済状況、3.生活における価値観・満足度、4.社会的行動の価値観・満足度、5.アイデンティティ、6.政治意識、7.健康状態のデータを蓄積したものである。

このアジアンパロメータを使ったデータベース活用法がユニークなのは、当該データの価値を十分理解している教員たち(代表は東京大学情報学環・東洋文化研究所園田茂人教授)が日本はもちろん、中国や台湾や韓国、それにアメリカ各地で講習会を開き、学生にデータの使い方を指南した上で、ワークショップ開催、コンソーシアム設立、コンテスト主催などを通じて、

研究者や学生を労力を惜しまず鼓舞して、優れた論文を世に送り続ける試みを行っていることである。

実際その研究成果[4],[5]は、日本の社会学会の中でも一定のポジションを得るにいたっていると聞き及んでいる。



図6. 東京大学東洋文化情報センター内
アジアンパロメータのフロントページ

しかしながら、すべてのデータベースにこのアジアンパロメータのような活用法が適用できるかという点、まず現実的に難しいことは容易に想像がつく。

第一に、このようなデータベースの活用には構築とは別の、ある意味では構築以上の「労力」が必要になることである。アジアンパロメータの例をみるに、主催する教員がこの所業を義務や負担といった後ろ向きの姿勢で行った時点で、試みは失敗する可能性を孕み始めることは明らかである。その問題がないとしても、年に何度も海外に向いて各種プランを練り直し、実施し、修正をする必要があろう。

第二に、その時その時に労力を厭わない「人材」の有無に大きく左右されざるを得ない、ということである。もしデータベース維持のために労力を厭わない人間がいても、もし当該データベースの周囲にその活用を意識的な人物がいなければ、休止するか忘れ去られるか、いずれかの事態に陥ることになるからである。そのような事態に陥るデータベースはどのような方法で休眠させて置くべきなのか、そしてまた時機が訪れたときにいかにして迅速それらを活用できるようにする体制にするべきなのか、問題はさらに根の深い問題に行き着くことになる。これについては本稿の最後にもう一度触れようと思う。

3. 継続活用を目指すデータベース周辺の技術変化

以上は、著者が実際に関わった、ないし関わっている機関におけるデータベースの具体的な活用事例である。

数少ない事例ではあるものの、その底流には次のような事実があるように思われる。

データベースというものは、その形態はあまり変わらないにもかかわらず、いつでも進化していく技術との比較によって古く感じられたり、新しく感じられたりするものである。例えば画像を使ったデータベースは、本来であれば画質の向上への対応につねにさらされる。数年前には高精細の最新カメラを使っていたものだったり、撮影方法を工夫して単体のカメラでは実現不可能だった画像を合成し格納していることを特長としたりしていたデータベースがあつという間に陳腐化してしまう例を個人的にも何度か目にしてきた。

また管理運用に関してはデータベースを格納するサーバ技術の変化も激しい。近年の象徴的な技術変化としては VM が挙げられよう。VM の急速な進化に伴い、クラウドという言葉もすっかり定着した。コストと安定性と情報保護とをトレードオフする選択肢が増えてきたといえるだろう。

しかし、安定した管理運用が活用のための基盤作業だということは現在も変わらない。サーバデーモン関連のトレンド、セキュリティ保持、インシデント情報の収集と迅速な対応体制を保持しておくことは、現在もこれからも必要なことであろう。

ただ、サーバ関連の技術が進化ないし変容しても、やはりそこでも相変わらず時限とコストの問題は関わってくる。この問題を解消するために現在は、活用に活路を見出すか技術に活路を見出すかの傾向は今後も続くはずである。

その手法としては、上記3つの個人的に内実を知りうる事例とは別に、最後に挙げた複数のデータベースを効率的に連携させる試み(データマインニングによるデータベース研究など)などが現在では代表的なものかもしれない。この手法は、連携した複数のデータベースを利用することによって、かえって、データベースの必要性を浮かび上がらせ、運用停止や管理保持へのカンフル剤の役割を果たす可能性をもつ。ただ、この手法では一つ二つのデータベースが停止してしまっても仕方がないという前提も含まれるかもしれない。また、例えば「デジタルアーカイブのコンテンツが地域の製品のブランド化と結びつく」といったより実践的な試みによってデータベースの維持を促すといった事例が、総務省が昨年にも出した「知のデジタルアーカイ

ブ社会の知識インフラの拡充に向けて「提言及びデジタルアーカイブの構築・連携のためのガイドライン」[6]に挙げられていたりする。が、しかしこれは学術データベースというデータベース領域には適切な例とはいえないかもしれない。

とまれ、これらもまた、データベースの内容のみを担保として維持管理していただくだけではなく、活用方法の工夫こそがデータベースの継続に寄与するというスタンスかと思われる。

4. データベース活用の別の視点

本稿では、異なるスタンスでデータベースを活用する方法を提案してみようと思う。

それは次のような前提から始まる。データベースというものはいずれ開発・構築が終わり、運用も終わるものである。

本稿ではあえて活用しているから止められないというスタンスをとらない。またはデータの内容によっては止めてしまっても構わないというスタンスもとらない。データベースは少なくとも構築当初は、構想者ないし開発者が何かしらの価値を付与して半自動的に情報を関連付けるデータの集合体である。したがって、本来であれば止めてよいデータベースは一つもないはずである。それにも関わらず管理運用の維持はやはり難しい。なぜだろうか。おそらくデジタルデータベースというものはそもそもコストや時限の問題で運用をいずれ止めざるを得ないものなのである。

これを前提とした上で、「停止に追い込まれないようにするためには」ではなく、「停止に追い込まれてしまっても」という視点からデジタルデータベースの管理運用ならびに活用を考えることはできないだろうか。

データベースがもし

- 迅速かつ的確に、再生できる機構を備えている

ものであれば、データベースの構築、管理運用、継続活用に関して新たな知見が生まれるのではなかろうか、というのが本稿の趣旨である。

4.1 再生の観点から見たプログラムソースと仕様

仮に停止に追い込まれたデータベースを考えよう。何が残されている可能性が高いだろうか。おそらく、報告書や仕様書の類、データ本体、ソースコードであろう。あるいは当時の開発者や利用者の記憶も「残されたもの」といってよいかもしれない。

データスキーマについては報告書や仕様書が残されていれば、参照できる可能性が高い。停止していても保存媒体が壊れていなければ、形式に問わず、取り出すこともできるだろう。ソースコードはどうであろうか。これもデータスキーマと同様、保存媒体が読み込み可能ならば取り出すことは容易だ。だが、問題はソースコードが読解できるかどうかだ。マイナーな言語あるいは再生を夢見る者が読めない言語である可能性は大いに考えられる。こうしてみるとソースコードに関しては、取り出すことはそれほど難しいことではないにせよ、再生の観点からみると、優れた報告書や仕様書には価値の点で及ばないと本稿は考える。

ソースコードより報告書や仕様書の方が価値の点で上である、という主張はプログラマからすると少し違和感を感じるかも知れない。しかし、例えば E.レイモンド『伽藍とバザール』[7]中、ソフトウェア開発の教訓その9.として全く同じことが言われている。

9: 賢いデータ構造と間抜けなコードのほうが、その逆よりずっとまし。

またもやフレッド・ブルックス本の第 11 章から。「コードだけ見せてくれてデータ構造は見せてもらえなかったら、わたしはわけがわからぬままだろう。データ構造さえ見せてもらえれば、コードのほうはたぶんいらぬ。見るまでもなく明らかだから」

実はその理由は簡単なものである。例えばボタンアクション一つを考えればよい。利用者はボタンを押すことに一つのアクションを割り当てているわけだが、プログラムではそうはいかない。単純なケースを考えても、変数に蓄えられたデータを引数として再度配列化したり、エラーチェックをしたり、出力プログラムに引数としてそれを渡したり、といった内容を書いていることが多い。ひょっとすると、サブルーチン化した関数を使いつつ、考えうる限り最適な結果を、自分なりにエレガントな方法で導き出そうとしているかもしれない。これらをソースコードから逐一読み解くより、開発者が何をしたかったのか、それを適切に表現できる手段があれば、その方が効率的で方法としてもスマートなのは明らかである。

データベースの再生の場合でも、どうやら報告書や仕様書に頼るのが良さそうである。

そうなるとデータベースを再生には迅速さという観点に加え

- どんなプログラム言語にも対応できる、再生のための機構を備えている

ことが同時に重要なポイントとして挙げられるよう。

そして、可能な限りの確・迅速に、プログラム言語に縛られずデータベースを再生できる仕組みの鍵は報告書あるいは仕様書にあると考えられる。

5. 仕様にかえれ コンセプトマップの提案

単なる成果報告ではなく、再生を予め考慮した報告書ないし仕様書作りとはどのようなものを目指せばよいだろうか。

本稿ではそのために 4 つの構成要素からなるデータベースのコンセプトマップなるものを提案したい。

考えられる 4 つの構成要素とは以下のものである。

1. 状態推移図
2. 画面推移図
3. プログラムフロー図履歴
4. データスキーマ
5. (データ本体)

重要度から並べていったが、これまでの記述経緯も踏まえ、順序を逆に、4.,5.から触れていこう。

5.1 データスキーマ

当然ながらこれはデータベースの命ともいべきものである。ここにはデータベース構想者の思想が詰め込まれているといってもよい。繰り返しになるが、最近ではリレーショナル構造をもつソフトを使うことが多いので、このスキーマの出力(例えば sq197 仕様であれば、desc)はいとも簡単に取り出すことができよう。データベース開発者の中にはスキーマを報告書や仕様書に載せることに違和感を感じるという者もいるかもしれないが、データベーススキーマはそれだけでは意味をもたない単なる形式、枠組みである。もちろん、全データが報告書や仕様書、できれば別途電子媒体の中に残されていれば、当初の目的である確・迅速なデータベース再生に役立つだろうが、データスキーマは報告書や仕様書には必須である。プログラムとの関係からこれは外せない。

5.2 プログラムフロー図とその履歴

次に 3.プログラムフロー図の履歴を取り上げよう。

プログラムフロー図というのは、非難されることも多いが、それにもいくつかの記述方法があると思う(例えば

古典的なダイアグラムによるフロー図などが、次のようなものである。

プログラムソースはたいいていブロックごとに記述されている。ここでは、ソースに直接あるいは別レイヤで視覚的にブロックを書き入れて、ブロック同士の相関を描画したものをプログラムフロー図として想定している(実際に市販されている解説書でこの手の方法をとるケースも多い)。

さきほどプログラムソースコードよりも報告書・仕様書を、と述べたが、次に述べる状態偏移図、画面偏移図の補助手段として、歴代開発者、メンテナの思想を追尾するために、プログラムフロー図をその歴史とともに残していくのはデータベース再生にとって強力な補助手段となりうる。

プログラムフロー図とデータベーススキーマは、開発者、代々の管理運用者ごとにログとして残しておいた方がよい。プログラムであればこれが changelog として記録される場所だが、あえて冗長性を増やしてより視覚的に保存される方がよいだろう。これらが重層的に記録されていることは、停止から再生までの期間が長ければ長いほど、役立つ可能性も生まれてくると考えられる。とりわけプログラムフロー図の履歴は種類の異なる言語で書かれた場合にはデータベース自体の資産になりうる。

5.3 画面偏移図と状態偏移図

最後にこのコンセプトマップ、報告書や仕様書にとって欠いてはならない 1.状態偏移図と 2.画面偏移図を取り上げる。先にボタンの例を述べたが、状態偏移図も画面偏移図もユーザアクションとプログラムアクションの相関図である。状態偏移図は開発者目線の、画面偏移図は利用者目線のそれといってもよいかもしれない。とまれ、両者はいずれもいわばデータベースの機能を凝縮し、簡略化したものといえる。

まずは画面偏移図である。

これは高い確率でデータベースのマニュアルに付されることが多い。それゆえ、おそらく構築後に残される可能性も高い。ただ、画面偏移図を作成するとき、利用者目線と同時にメンテナ目線も考慮することが必要かもしれない。簡易化することは効果があるが、メンテナにとって必要な情報が落とされては何の役に立たないものになりかねないからである。報告書などを書く場合に、この部分に積極的に開発プログラマは関わるべきである。

次に状態偏移図である。

これも簡単なものだけに大規模な報告書には開発者が付す場合も多い。例えば図7.は拠点アーカイブ

PSS のための状態偏移図であるが、この図だけでは例えば Pompei/Rome が何を指しているのかがわからないが、もしこれが効果的に画面偏移図と連動していればその意味は自ずとわかるであろう。

さらに、これらに加えてデータスキーマがあれば、もっと具体的に多くのプログラマにコンテンツ再生を可能にするものと思われる。

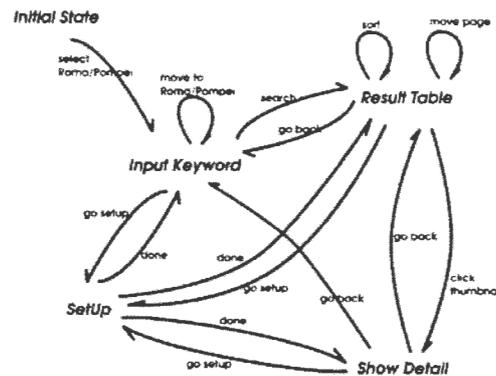


図7. 状態偏移図の例

重要なのは、ここに特定の言語の記法に関する知識は可能な限り排除されている点である。

図7.にも "select" といった一部 sql 文を想起させる部分があるが、それでもここで使われている機能は "move to", "sort", "search", "go to", "go back", "done", "click" などごく単純な用語で示されている。これらの機能の実装はそのつどのプログラマに委ねられている。

どのように実装してもよい、というスタンスでプログラムの記録が残されていることは、データベースを再生させようとする者にとって、時として自身のプログラム技術以上の効果をもたらす。すなわち、迅速なプロトタイプの実現である。そして、この者は実装が固まった時点でもういちど自分が手本とした状態偏移図を眺め、追加・変更した機能について修正を施す。

こうすることでデータベースは単に機能するもの以上に、「発掘可能な所産」となるのではないだろうか。

5.4 コンセプトマップ

以上 1.から 4.の要素を効果的に関連付けたものを本稿では仮にコンセプトマップ(参照:図8)と呼ぶことにする。

仮に、といったのはこれがまだ構想段階であるからで、実際に運用中ならびに停止してしまったデータベースに果たして本稿のアイデアを実際に検分する必要があるからである。

ただ構造がシンプル極まりないだけに検分手順も大方見通しがついている。

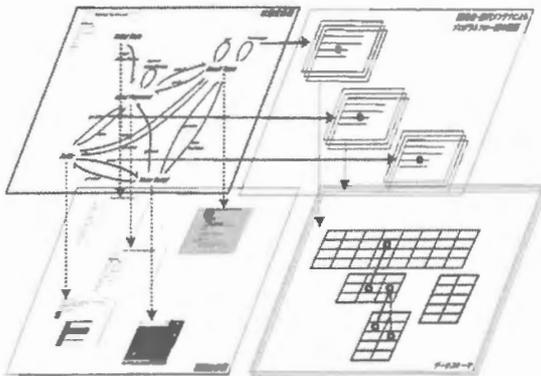


図8. コンセプトマップ(案)

1. 「運用中の」小規模なデータベースを素材にコンセプトマップを開発者に依頼。マップ作成にはどれほどの手間がかかるかを実測。改善点を検討。
2. 1.を別言語で再開発。構築時間の実測。改善点の検討。
3. 1.を今度は「停止している」が当時の開発者と連絡がとれる小規模なデータベースで実施。
4. 3.を 2.を踏襲して実施。元開発者に検証を依頼。
5. 1.から 4.の成果を整理し、「開発者とのコンタクトもなく、運用も停止している」データベースにて再生を実施、検証。

データベースの構造は今後も進化していくべきである。コンセプトマップは、データベースを再生しようと目論む者のアイデアを、できるだけ迅速に、そして実行に移す際の心的負担を少なくして、反映させるという目論見を元に構想されたものである。

6. コンセプトマップから、データベースのバザール型形成の可能性を探る

実はこのコンセプトマップ構想にはもうひとつの可能性を込めている。それはデータベースの活用形態の新たなオプションとしてのバザール型コミュニティ形成である。

コンセプトマップ構想の実現には、運用を停止したデータベースを対象としているにも関わらず、多くの手間と時間がかかるだろう。

だが、このこと自体がデータベースを「活用」していることにはならないだろうか。

コンセプトマップはデータの内容を含むものではない。よって、著作権等には抵触する確率が低い。これを特定のデータベース固有の資料として保存することには

ほとんど意味がない。それでは、もしこれをデータベースの構造に関心をもつ研究者が共有したらどうだろうか。共有化の結果、構造の類型化がなされ、それを参考に管理運用中のデータベースの改善点や機能追加のヒントにはならないだろうか。あるいは停止してしまったデータベースの「発掘」の重要な緒にはならないだろうか。

従来のデータベースの活用のプレーヤといえば、無自覚にその利用者を指してきた。換言すれば、データベースの活用方法というのは、データベースの利用法と同義であった。だが、このコンセプトマップ構想は、開発者、管理運用者、将来のデータベース開発者を巻き込む。しかも、将来のデータベース開発者に、複数のコンセプトマップを提示できれば、より効率のもしくはより斬新なアイデアの種を提供できる可能性すらあると考えられる。

繰り返しになるが、コンセプトマップ構想ないしコンセプトマップ仕様の策定に見込んでおいてよいことは

- データベース活用のフィールドに新たなプレーヤが自然と加わる

ことである。

というのも、この手法にある程度の見通しが立てられた暁には、先に挙げた E.レイモンドが掲げた「バザール型」開発と同じ環境をデータベース開発者・再生者が得るからである。

この開発形態の優位性はすでに広く認められている。その特長として頻繁な更新回数と対象(ソフトウェア)の譲渡のスムーズさが挙げられる。

運用停止の危機に陥っている、ないしはすでに停止してしまったデータベースにとって、さらには今後の停止の危機に備えるためにも、データベースのコンセプトマップ公開、収集はマイルストーンになりうる。

なるほどデータベースを最初に構築する際は、確かにデータに関する監修者が必要で、そのニーズに応えることが必須条件であるが、その後はプログラマがプログラムを共同で開発する形態にそれを持ち込むことができるのである。

データベースを活用することと、必要に応じて再生させることは、今後のデータベース開発、管理運用に多少なりとも貢献できるアイデアではないかと本稿は期待するものである。

参考文献

- [1] 東京大学大学院人文社会学系研究科象形文化研究拠点, 「文部科学省中核的研究拠点 COE 形成プログラム報告書象形文化の継承と創成に関する研究」, 2004.
- [2] 神内俊郎・浜田長晴・池庄司伸夫, 「デジタルイメージシステム (DIS) の開発とその応用」, 日立評論 デジタルアーカイブ, 1997.
- [3] 国立西洋美術館, 「古代ローマ帝国の遺産 栄光の都ローマと悲劇の街ポンペイ」, 2009
- [4] 園田茂人編, 『勃興する東アジアの中産階級: アジア比較社会研究のフロンティア』勁草書房, 勁草書房, 2012
- [5] 園田茂人編, 『リスクの中の東アジア: アジア比較社会研究のフロンティア』勁草書房, 2012
- [6] 総務省, 「知のデジタルアーカイブ ー社会の知識インフラの拡充に向けて ー提言及びデジタルアーカイブの構築・連携のためのガイドライン」, <http://current.ndl.go.jp/node/20515>, 2012
- [7] Eric Steven Raymond 著, 山形浩生訳, 『伽藍とバザール』, 光芒社, 1999