

活断層データベース改訂版のデータモデル

A data model for reconstruction of Active Fault Database in Japan

伏島 祐一郎 [1]; 三輪 敦志 [2]; 宮本 富士香 [1]; 吉岡 敏和 [1]
Yuichiro Fusejima[1]; atsushi Miwa[2]; Fujika Miyamoto[1]; Toshikazu Yoshioka[1]

[1] 産総研 活断層研究センター; [2] 応用地質
[1] Active Fault Research Center, AIST, GSJ; [2] OYO

<http://unit.aist.go.jp/actfault/activef.html>

活断層データベースは、産業技術総合研究所活断層研究センターによって構築され、2005年3月にインターネット上に公開された (<http://www.aist.go.jp/RIODB/activefault/>)。このデータベースは、リレーショナルデータベースマネジメントシステム上で運用されているものの、データ入力を表計算ソフトウェアによって行ったため、リレーショナルデータベースとしての真価を発揮できずにいる。すなわち、限られた表示様式による電子閲覧機能のみに特化した、電子カタログの段階に留まっており、検索・並べ替え・分析・描画・自動制御など、リレーショナルデータベース本来の旨味を生かす道が閉ざされている。そこで真のリレーショナルデータベースへの再構築を目指し、根本的な改訂作業を進める事となった。この改訂作業の基礎となるデータモデルと、その構築過程について紹介する。なお大会当日までに、このデータモデルに基づくデータベース改訂版が、公開される予定となっている。

データモデルとは、リレーショナルデータベースへのデータ格納規則とその設計図であり、博物館収蔵庫の配架規則に喩えるのが最も理解しやすいだろう。どの収蔵品(データ)を、棚(データベーステーブル)のどの段(データ項目)に格納するか; それぞれの段の大きさ・種類をどう規格化するか; そして複数の棚をどう並べ関連づけるのか(リレーションシップ); それらは詳細かつ明確な規則に従う必要がある。この規則は、単に個々の収蔵品を管理するという目的だけではなく、多様な収蔵品を組み合わせた多様な企画展示を開催する事に目的を置く。その為に収蔵品は、個々の展示に支配されない独立した立場を確保する必要があり、この事が、リレーショナルデータベースにとって最も重要な、データの独立性確保の比喩となりうるだろう。

すなわちデータモデルは、ソフトウェア・ハードウェア・利用者・利用目的等々の多様性を超えた、データそのものによる論理モデルであり、それによって多様な検索・並べ替え・分析・描画・自動制御が実現されるのである。リレーショナルデータベースにおいては、この規則は集合論や論理演算など、明快で美しい理論の上に構築されている。それらに基づきリレーショナル演算を複数のテーブルに施す事によって、無限の表による無限の表示結果を導き出し、さらに分析・描画・自動制御の為に多様な構成のデータを供給する事ができるのである。

データモデルの構築過程を以下に略記する: まずデータモデルへの変換をおこなうべく、研究過程の理想モデルを作成した。このモデルは、研究 データ生成 データベースへの入力 出力 研究という、データベース利用サイクルの上に構築する事が重要である。次に、データ量・データ利用頻度・予算・構築日程等の現実的条件を勘案・妥協し、モデルから研究過程を取捨選択した。そして選択された研究過程をグルーピングし、プロセスモデルを作成した: すなわち調査研究 調査地 変位基準・変位・イベント 活動性評価 文献 調査研究というサイクルにモデル化した。さらにこのプロセスモデルを、量的対応関係等の、データの存否に関するモデルへと変換する事によって、データモデルの原型を作った。

それぞれの研究過程において生成されるデータ項目を、勘案と妥協の上で抽出し、テーブルの各列として定義することによって、データモデルを明確化していった。次に各データ項目を可能な限り規格化し、それをマスターテーブル(凡例表・コード番号表)に置き換え関係づけた=すなわちリレーションを張った。これによりデータのかなりの部分は、数値に置き換えられた。そしてデータの軽量化や、一括修正等のデータ管理機能強化、集計等の分析の基礎が築かれた。さらにこのデータモデル構築に於いては、地質学に特徴的な曖昧数値データを、計算・分析に利用可能な明確な数値データに変換する試みを行い、作成した数値データテーブルとのリレーションを張っていった。

以上によるテーブル作成を終えた後に、ループするリレーションシップを置き換える等の、論理的改善作業を行い、データモデルをコンピュータに馴染ませた。最後にそれぞれのテーブルについて、そのデータ項目の型・文字数等を詳細に定義し、データモデル構築を終えた。定義されたテーブルの総数は、50を超えた。

実際の構築作業は、妥協の積み重ねとデータの細かな仕様定義に明け暮れるものであった。また、これが地質学・地震学分野ではほぼ日本初の本格的データモデル構築となった為、手探りで解決しなければならない問題が山積していた。とは言え大会では、泥臭い作業の思い出話は避け、データモデルを具体的に示すER図を掲げるとともに、比喩によってデータモデル構築過程のイメージをつかんでいただくよう心がけた。