

2Q-2

サイエンティフィックデータベースのための一次データの一管理法

中田 充 寶珍輝尚 都司達夫

福井大学 工学部 情報工学科

1はじめに

実験データや測定データなどの科学技術データは、必ずしもデータベース化を前提として生成されているわけではない。従って、データベース化の際にデータ定義（スキーマ）を決定することが困難である。さらに、これらのデータのスキーマは、研究者の視点や仮定によって決定されるため、それらのコンセンサスが得られるまでは、多くの研究者が納得できるスキーマを決定できない。このことから、科学技術データをデータベース化するには、データをとりあえず未整理の状態で格納する一次データベースと、十分な検討や試行錯誤をくり返しながらスキーマを決定していく機構が必要である。

我々は、研究者の仮定や視点に依存するデータ値の格納や、データの複数の解釈の柔軟な表現、を可能とするためのデータモデル[1]を提案しており、それに基づいたDBMSであるDREAM (Database management system for stoRing the sciEntific data and Adding new Meaning) を設計・作成している[2]。DREAMは、一次データベースである基本データベースとそのデータを参照する導出データベースの2種類のデータベースを管理する。導出データベースは、研究者の視点や仮定に沿って基本データベース中のデータを参照する以外に、各研究者が独自に生成したデータを他の研究者から独立して管理する役割を持つ。本稿では、基本データベースの管

理法について述べる。

2 基本データベースに対する要求**1) 未整理データの格納**

格納するデータは、未整理データであるので、データ定義を必要としないデータ格納機能が必要である。

2) 測定条件の管理

科学技術データにとって、データを得た環境や測定の条件は重要である。従って、データを得た時の条件を管理する機能が必要である。

3) 同一対象からのデータのグループ化機構

繰り返し行われた測定や実験によって得られた、同一対象に対する複数のデータをデータベース化する場合、データ毎に識別子を割り当てる、同一の対象物から得られたデータであることを認識することが難しくなる。従って、同一対象から得られたデータをまとめる機能が必要である。

3 基本データベースの設計

要求1)を満足するため、データは基本的に文字列データとして格納することとする。画像などのマルチメディアデータは、名前を付加したバイナリデータとして管理する。さらに、これらのデータには、それを得た環境や測定の条件を表わす測定条件を付加する。これにより要求2)を満たす。このデータと測定条件の組を基本データと呼ぶ。基本データには、一つのテキストデータと0個以上のバイナリデータが含まれる。図1は、考古学上の遺跡から発掘された陶器の破片に対する基本データの例である。図の基本データは、テキストデータとして破片表面に描かれている紋様の名前を含み、バイナリデ

Management of Primary Data for a Scientific Database

Mitsuru NAKATA Teruhisa HOCHIN

Tatsuo TSUJI

Dept. of Information Science, Fukui University

ータとしてそれらのイメージを含む。バイナリデータに付けられた名前（“外側紋様の画像”）をテキストデータ中に含めることで、バイナリデータの存在を表している。基本データの測定条件として、測定者名、測定日時、測定に使用した器具などの記述がある。要求3)に対応するために、一つの識別子と基本データの集合との組を基本エレメントとする。基本エレメントは複数の基本データを含むことが可能であり、ユーザは、各基本データの測定条件をもとに所望するデータを決定する。基本データベースは基本エレメントの集合として定義される。

4 基本データベース管理システムの実現

ワークステーション HP9000S730 上で、商用の OODBMS UniSQL を使用してプロトタイプを構築した。プロトタイプでは、基本エレメントを格納するための UniSQL のクラスを定義し、そのクラスを用いてデータを管理している。クラス定義を図2に示す。basic_data は基本データを表わすクラスであり、測定条件を表わすクラス measure_cond と測定データを表わすクラス measure_data をインスタンスとして持つ。クラス measure_cond は、測定条件を文字列で格納する。クラス measure_data は、データを文字列で格納し、複数のバイナリデータを持ち得る。クラス multimedia は、バイナリデータを扱うクラスであるが、現在は、X11bitmap の画像データのみを扱う。

5 基本データベースの利用

基本データベースに格納された基本データを、導出データベースから参照し、研究者の視点に沿ったデータを構築する。基本データを導出データベースから参照するために、指示エレメントを利用する。指示エレメントは、流動的データ型[3]を用いて実現しており、基本データの一部分を仮想データとして参照することを可能にしている。また、指示エレメントには、データ

型変換の機能があり、文字列データとして格納されているデータも、導出データベースにおいて、数値データとして参照することができる。

6 おわりに

未整理データを格納するための一次データベースの管理法について述べた。現在、福井県一乗谷朝倉氏遺跡より発掘された陶器の破片データを基本データベースに格納し、導出データベースから参照できる状態である。基本データベースのユーザインターフェース、効率の良い検索機構が今後の課題である。

参考文献

- 中田充、宝珍輝尚、都司達夫：サイエンティフィックデータベースのためのデータモデルの一提案、情処学会研究報告、95-DBS-101-9, (1995).
- 中田充、宝珍輝尚、都司達夫：サイエンティフィックデータベース管理システム DREAM の設計、情処学平成7年後期全大、5E-09, (1995).
- Hochin, T. and Tsuji, T. : On the Application Interface of the LIQUID Data Type for Flexible Manipulation of Multimedia Data, Proc. of ADT'94, pp.200-206 (1994).

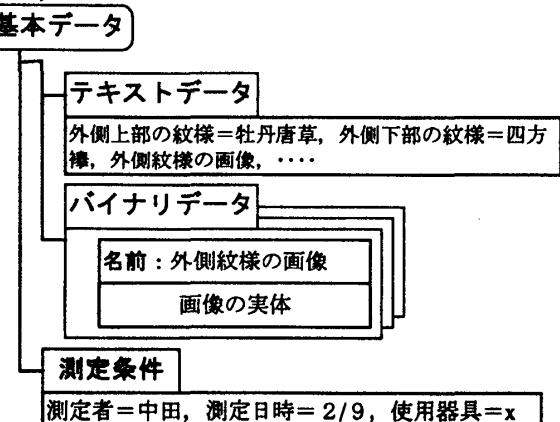


図1 基本データ

```

create class basic_data (mcond measure_cond,
                        mdata measure_data);
create class measure_cond (condition string);
create class measure_data (fact_data string,
                           bulk_data sequence_of multimedia);
create class multimedia (img_name string,
                        img_data set_of (x11bitmap, glo));
  
```

図2 クラス定義