

解説

データ・ディクショナリ/ディレクトリ†



椿 正 明††

1. はじめに

データ・ディクショナリ/ディレクトリ(以下 DD/D という)は、情報システムの核として、急速に注目を集めつつあるが、かつての DBMS に比べると、それもよほど落付いたものにみえる。パッケージの価格が安いからというよりも、そのご利益が間接的かつ長期的で、大衆向けにいく、したがってメインフレームの売込みにも直結しにくいからであろう。

DD/D もシステム開発・運用のツールではあるが、DBMS が半製品であるのに反して、最終製品である。情報システムそのものを対象とし、データ管理者、データベース管理者、プログラマを主なユーザとする応用システムの一つである。時々そのデータベース(あるいはファイル部分)のみを指していわれることもあるが、一般には、データベースとこれにかかわるプログラム一式から成るソフトウェアを指す。その意味では、データ・ディクショナリ・システムあるいは DDS と呼ぶ方が妥当ともいえる。

DD/D は、しばしば「メタデータを蓄える」といわれる。これは、DD/D の蓄える最も代表的なデータが、社員番号、氏名、所属、基本給といったいわゆるファイルの項目(あるいはフィールド)に関するデータだからである。このファイルの項目は、DD/D の世界ではデータ・エレメントと呼ばれることが多い。

DD/D への関心が強まってきた第一の原因は、情報システムの大型化、複雑化により、扱うデータエレメントの数が増加したことである。DBMS を適用したりして大規模システムを構築していったとき、システム関係者自身、その扱うシステム部品をコンピュータで管理する必要を感ずるに至ったわけである。したがって DD/D パッケージが商品化されたのは、実用的なデータベース・システムが広く実働化され始めた 1974 年ころからである。

† Database Consultant by Masaaki TSUBAKI (Nihon Systemix Corporation).

†† (株)日本システムックス

DD/D に関する研究論文は、まだきわめて少ない。データベースの場合と同様に、まず商用パッケージが先行し、その研究は、これを追いかけるように、これから進められるものと期待される。しかしその例外とでもいうべきものであろうか、1974 年に発表された GUIDE の Data Dictionary/Directory に対する要望書 (GUIDE 74)、および 1975 年の ANSI/X3/SPARC の 3 層スキーマ構造を提案する報告書 (ANSI 75) は忘れることができない。前者は、きわめて早い時期の DD/D に関する要求仕様にもかかわらず、その内容は、エンティティ概念にもとづくきわめて本格的なものであり、昨今の意味論データモデルの立場から見ても決して古さが感じられない。一方後者は概念スキーマ、外部スキーマ、内部スキーマ相互の変換を司る中核として、情報システムにおける DD/D の位置づけを明らかにした点で大きな功績を残したといえる。ただし、時代の制約によるとはいえ、両者とも DD/D に DBMS の支援ツールとしての意義しか認めていない点に注意しておこう。

情報システムを構成する部品タイプとして、データ・エレメントが最も代表的なものであることに異論のあるはずもないが、このほかレコード、ファイル、プログラム、モジュールなど種々の部品タイプが、構成部品タイプとして挙げられる。しかもこれらは相互に種々の関連を持っている。したがって、これを表現するために、DBMS の支援ツールである DD/D が逆に DBMS の支援によって構築されることも、きわめて自然の成行きといえる。こうして、DD/D は、DBMS を駆動するもの (Active)、しないもの (Passive) のほかに、DBMS に依存するもの (Dependent)、しないもの (Independent) として分類される。

DBMS は、それ単独としてよりも、入力処理、レポート・ライター等種々の機能と連繋して、より強力な支援効果を上げる。そして、そのことごとくが、DD/D 情報を必要とする。そこで機能別に個々に用意されていた DD/D を統合集中し、DBMS をはじめ関連する

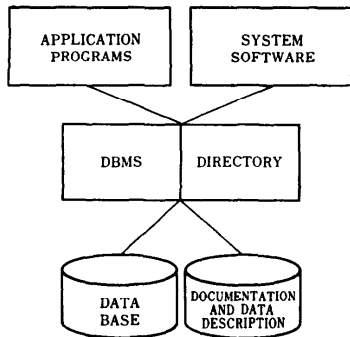


図-1 R. Weeks の提案する IDD の構成

プログラムがこれを共用するいわゆるファミリー化パッケージが、昨今の流行となりつつある。このような DD/D は、しばしば IDD (統合化 DD/D) と呼ばれる。IDD 化の最大のねらいは、メタデータの冗長性をなくすことによって、運用を合理化し、質を向上させることと、整合性の高いパッケージを提供しようとするものであるが、DBMS ベンダにおいては、同時に、DD/D と DBMS の融合を志向している。Cincom 社の R. Weeks によるとそのイメージは、図-1 のように表わされる。

IDD 化とともに DD/D を単なるデータ・エレメントの管理から脱皮して、情報システム全般の管理を志向する情報システム資源管理ツールとして位置づけるのが著しい傾向である。

以下その現状と解決すべき問題点を中心に解説を試みることにしよう。

2. 商用パッケージ

商用パッケージとしては、表-1 に挙げるものが代表的であろう。はじめの3つは、そのファイルの構築、

したがってアクセス手段を汎用の DBMS パッケージに依存しないもの (Independent)、後の6つは、それぞれ DBMS に依存するもの (Dependent) である。前者の場合は、アクセス手段の影響を受けないため、いくつかのアクセス手段を用いるヘテロジニアスな環境、いいかえると EDP 化の歴史の古い先進企業などにも適用しやすく、適用範囲が広い。一方後者の場合は一般に歴史的に特定の DBMS 下のアプリケーションの主としてデータ・エレメント、レコード、ファイルの管理、DDL の生成などに重点があったため、その影響が強く、その DBMS を用いて全システムをまとめる場合は良いが、ともすると、その DBMS の係わるシステムだけをつまみ食いの的に支援し、組織全体のデータをグローバルに見る視点を失わせやすいので注意が必要である。

前述のように、DD/D の IDD 化が、流行となっているが、個々のパッケージにおけるその達成度を判定することはきわめて難しい。広告で見限る限り、表-1の後の4つが IDD を強調しているが、メタデータの冗長度の定量的評価は困難であり、またファミリー化パッケージの数で IDD 化の達成度をみるわけにもいかない。メタデータを IDD に統合集中し、完全に冗長性をなくすことは現在のところ現実的でない。しかしこれをせめて一旦はまず、IDD に登録し、次にこれを各用途に用いるべく配布すべきであるが、DDL やソース、プログラムを解析してメタデータを集めたり、いわば「後向き」の DD/D 機能をサポートしているものが多い。

しかしどのパッケージについてもいえることは、IDD 化、あるいはトータルな情報資源管理をめざして、確実に機能向上を続けていることである。主たる DBMS に加えて順ファイル、VSAM、IMS などポピ

表-1 代表的な商用 DD/D パッケージ

パッケージ名	開発者名	依存 DBMS	支援対象
IRM (PRIDE-LOGIK)	MBA 社	独自	全般
DATAMANAGER	MSP 社	独自	全般
DATA CATALOGUE 2	Synergetic 社	独自	全般
UCC TEN	UCC 社	IMS	主に IMS アプリケーション
DB/DC DD	IBM 社	IMS	主に IMS アプリケーション
S2000/80 IDD	Intel 社	S2000/80	主に S2000 アプリケーション
IDMS IDD	Cullinane Database Systems 社	IDMS	主に IDMS アプリケーション
DATADICIONARY	ADR 社	DATACOM/DB	主に DATACOM/DB アプリケーション
シリーズ 80 DCS	Cincom 社	TOTAL V.8	主に TOTAL アプリケーション

ュなアクセス法をサポートするとともにレポート・ライター、スクリーン・サポート、テストデータ生成、プログラム生成などの機能を追加している。またメタデータの範囲もシステム、プログラム、モジュール、要員、組織などを追加し拡大させている。

表-1には挙げなかったが、今や DBMS にとって DD/D は不可欠となりつつあり、ADABAS, MODEL 204, INQUIRE, ORACLE なども DD/D を充実させているが、これらについてもほぼ同じ傾向が見られる。総じて各パッケージとも、他の長所をとり入れつつ、同じ方向に発展しつつあるかに見える。

3. DD/D の用途

現状を正しく把握することは、あらゆる行動の原点である。情報システムの構造を正しく記述した DD/D は、予想を超えて種々の用途を生み出しつつある。主なものを挙げると次のとおりである。

- (1) システム運用・保守支援
- (2) システム開発支援
- (3) DBMS 駆動
- (4) 分散 DB の仮想化
- (5) 異種システムの統合
- (6) DBMS 独立の達成
- (7) 標準化
- (8) 情報システム計画支援

(1)は変更の際際のインパクト分析など、最も古くからの用途である。システム開発が済んでからドキュメントを整理しておく、いわゆる「後向きのドキュメンテーション」をコンピュータ上で実現してもこの用途としては間に合うからである。

(2)は、システム開発の途上でもっと「前向き」に DD/D を利用するものである。すなわち DD/D をエンジニアリング・データベースとしてとらえ、システム部品が識別され、仕様が決定される都度、これを DD/D に投入してゆく。この場合、データ・エレメントに限らずあらゆるシステム部品が記述される。大勢の開発担当者は、これを共用しこの上でコミュニケーションすることができる。実際 W. Kent (KENT 81) も、初め DD/D をデータ記述のツールとして用いていたが、データ設計にきわめて有効であったと述べている。

DDL の生成、COPYLIB の生成、プログラムの生成、テストデータの生成、レポート・ライター支援などは、このような DD/D の利用の機械化版とみること

ができる。DD/D の「前向き」の利用は、決して容易ではないが、機械化は、これを促進する上で効果があるものと期待される。

(3)は、DBMS が、検索・更新等のサービスを遂行するに当たり、DD/D を利用するものである。筆者はかつて自己記述型の DBMS, DPLS においてこれを実現したが最近の DBMS では、この方式をとるのが増えつつあるようだ。OS 駆動もこの延長線上にある。

(4)は、ユーザにデータベースの分散を意識させないための DD/D の利用である。DD/D にはデータベースの所在が記述されている。システムはユーザ要求にこたえるために、どこどのデータベースにアクセスすべきかを DD/D を参照して判断する。こうしてより高度の情報隠ぺいを実現される。

(5)は、異種アクセス法によって構築された複数個のシステムを一個のシステムとして統合するものである。英和辞典が異種言語の変換を司るように DD/D が異種システム間で必要とされる変換を実現する。異種ネットワーク間の通信等もこれと同列に考えることができよう。

(6)は、概念スキーマと DBMS 論理スキーマを DD/D に記述し、この間の変換を司るデータベース・アクセス・モジュールを介して応用プログラムがデータベースにアクセスできるようにするものである。こうして応用プログラムは、データベースに対して概念スキーマ・インタフェースを持つことになり、DBMS に関するポータビリティを獲得する。すなわち通常のデータ独立性よりさらに次元の高い独立性「DBMS 独立性」が達成できる。

(7)は、DD/D の活用によって自然に達成できる効果である。システム部品の仕様を、DD/D に明示的に記述することによって、システム関係者同志あるいはエンドユーザとの間のコミュニケーションが正しくかつ効率的に行われ、標準化が進む。データの重複や矛盾が明示されるのでこれが自然に排除されることになる。

(8)は、比較的最近注目されつつある用途である。IBM の BSP (Business system planning) が有名であるが、企業等の情報システム全体像をトップダウンにとらえ、長期計画を立てる ISP (情報システム計画法) に、DD/D を適用しようとするものである。企業等を丸ごと情報システムと見立て、その組織設計等に利用しようとする、いわば企業駆動としてその用途も

注目されつつある。

4. DD/D の位置づけ

DD/D が、当初特定の DBMS アプリケーションに対する支援ツールあるいはシステム開発のドキュメンテーション・ツールとして生れた頃は、単に1つの開発者向け応用システムとして考えればよかった。しかしその用途が広がり、情報システムの中核として、他システムと密接な関係を持つようになると、その位置づけを明確にしておく必要が生れる。

たとえば次のような疑問にどう答えればよいであろうか。

(1) DD/D データベースは、他の応用システム・データベースとどういう関係にあるのか。

(2) DD/D データベースとは、1企業に1個あるべきものか、 n 個あるものか。

(1) に関しては、「DD/D は1個のデータベースの中の一部である」とするものから、「企業内にある n 個のデータベースのうち特殊な、あるいは普通の、1個あるいは m 個である」まで種々あり得るだろう。したがってこれに答えるためには、情報システム環境をマルチ・データベースで考えるのか、またデータベース1個をどう定義するかなど、を明確にしておかなければならない。

(2) に関しては、情報システムを現に駆動している確定的データと、新規開発あるいは改訂中のデータとをいかに区分するか、また分散拠点における DD/D の重複配置をどう扱うかが問題になる。

以上のように DD/D の位置づけを明らかにするには ANSI/X3/SPARC とは別の意味で、データベース・アーキテクチャをどう設定するか、その枠組を決定しなければならない。しかしこのような議論はまだほとんど行われていないのが実情である。私見を述べるならば、分散マルチ・データベース環境の中に、運用および開発用の DD/D を区別して、図-2 のように位置づけて考えたい。ここで各分散拠点における運用

DD/D は、原則として同じ内容を持ち、特別な時間帯に同時に更新されるほかは参照だけが許されるものである。なお応用 A データベースは、同型式ではあるが内容を異にする、いわゆる分散データベースを形成するもの、また応用 C データベースは、同型式で内容を異にするものが同一拠点に2個生成されたものを表わしている。

5. 解決すべき課題

DD/D の位置づけもその1つであるが、商品化を軸に試行錯誤的に発達してきた DD/D 技術には、解決すべき問題が山積している。R. M. Curtice (CURT 81 b) はこれについて、次のように述べている。

(1) 良いデータ定義を構成する規準はおろかガイドラインすらない。

(2) DD/D にデータのどのような属性を記録しておくべきか明らかでない。

(3) データ・エレメントの有効な定義がない。

(4) 概念スキーマあるいは論理スキーマの設計法が確立していない。

(5) 概念スキーマ・モデルについて議論が紛糾している。

DD/D は、要するに情報システムを部品に展開し、個々の部品の仕様と相互関係を記述するものであるが、どのような部品タイプに展開すれば良いか、まだ結論が出ていないわけである。いいかえると、情報システムに対するデータベース設計が完成していないまま、それぞれ部分的に実用化されているのである。

したがって、各パッケージによって識別されるエンティティ・タイプはマチマチであり、異種 DD/D 間のコミュニケーションはきわめて難しい状況にある。標準化のツールが逆に標準化を妨げるという矛盾をはらんでいるわけである。

多くの問題のうち、研究者に委ねられた、とくに緊急かつ重大と筆者が考える課題は次の2つである。

(1) データ・エレメントの定義を、定義域、アトリビュートと関係づけて明確にすること、すなわち概念スキーマを構成する3大要素は、図-3 のように定義域、アトリビュート、概念ファイル(ファイルを論理化し、かつローカル・ビューを統合化したもの、一般にエンティティ・タイプと呼ばれるものと同一視できる)から成ると考えられるが、データ・エレメントは、定義域をいうのか、アトリビュートをいうのか、そのいずれとも異なるのかを明らかにしなければならない。

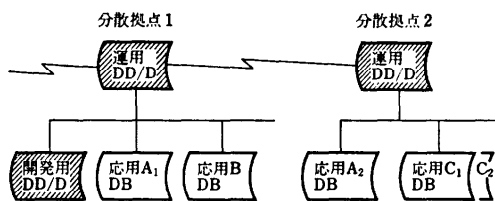


図-2 DD/D の位置づけ

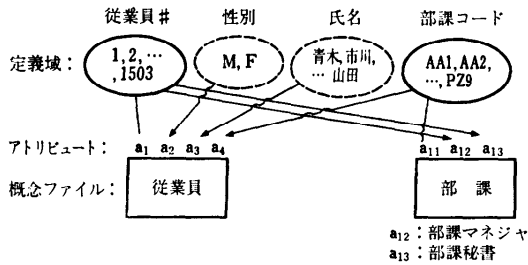


図-3 定義域, アトリビュート, 概念ファイルの関係

(2) 企業トータルシステムあるいは企業間システムなど広範な情報システムに対処できるよう概念スキーマを位置づけ, そのデータモデルを確立すること, この場合外部スキーマとの相違を DD/D 上での記述によって明示し, W. Kent(KENT 80)の Collective Schema などの問題を解決すること。

現在 DD/D パッケージは, 唯一の例外 (IRM) を除いて, システム設計・開発のための方法論ぬきで販売されている。システム部品タイプをユーザが自由に定義できるようにしたものも多い。使い方に関してユーザに大きな自由度が与えられているわけである。ユーザは上手く使う可能性と同時に, 下手に使い, さらに失敗する危険性も多く与えられているわけである。

最も警戒すべきことは, DD/D の内容が実体と食い違うことである。これは情報システム全体に大きな被害をもたらす可能性がある。このようなことが時々でもあると, DD/D が信用されなくなり, まますます汚染がひどくなる原因を作る。これを防ぐ決め手は, やはり, システム開発の方法論を確立しデータ管理体制を整え, 正しい手順で DD/D を使用することである。正しい手順とは, 設計作業の中でシステム部品を識別したら, それをその都度直ちに DD/D に投入してゆくことである。設計ドキュメントは, DD/D 出力として作ることである。ユーザとの, あるいは設計者同志のコミュニケーションは, DD/D を介して行うことである。ドキュメントはドキュメント, DD/D は DD/D と二本立てで運用すると, DD/D はどうしても「後向き」のものとなり, 品質低下を招くことになりやすい。方法論がまずあってこれを支援することを目的として開発された DD/D はこの点を重視しているわけである。プロジェクト管理との連繋もこのような DD/D によってはじめて可能となる。

6. DD/D の研究

DD/D に関する研究は, 前述のとおり, まだ未熟な段階にある。データベースに関する研究の中心課題であったデータ・モデル論は, 本来 DD/D の研究課題として考えた方が良かったものも多いが, そのような意識では行われなかった。

リレーショナル・モデルでは, CODASYL モデルでは, あるいは E-R モデルでは, 一体 DD/D の中にどのようなエンティティタイプを定義し, どのような対応関係を登録することになるのかを述べれば良かったのではないか。それによって, 冗長な説明や難しい数式の大半を省略し, かつ無用の誤解を避けることができたのではないか。また DD/D の構造と比較すれば, DDL のシンタックスを決めることなどおおよそ本質的でないことが容易に見通せたことと思われる。

DD/D に関する研究は, 従来主に DD/D ベンダや DBMS ベンダによって, 実用的な観点から進められてきたわけであるが, アカデミックな研究がもっとあってもよいと思われる。情報システムは研究者にとっても十分身近に接することができ, また, メタデータに関する抽象的な思考は, むしろその得意とする部分のはずだからである。

少ない中ではあるが, 前述した R. M. Curtice の論文 (CURT 81 a), (CURT 81 b) は出色である。的確な問題提起がきわめて貴重である。W. Kent (KENT 81) の Sharade (DD/D) 使用の経験も興味深い。また統計データベースの分野で DD/D についての関心が深まりつつある (たとえば (CHAN 81))。

DD/D にもとづいたプログラム生成の研究も始まったようである (たとえば (RISC 80), (TAMI 80))。入出力の仕様と導出データの条件を与えてプログラムを生成する方式は, プログラム開発合理化の最も基本的なアプローチと思われる。この辺でソフトウェア工学との融合が成立すると面白い。

わが国では池田らを中心とした広大グループの実践にもとづいた研究 (たとえば (IKED 81)) が注目される。今後のより深い理論的展開を期待したい。

なお研究というよりも, DD/D を解説した資料として, (BRIT 77), (DATA 78), (DBSE 76), (DIEB 79), (LEFK 77), (UHRO 73) などが参考になろう。

7. おわりに

DD/D は情報システム資源管理ツールとして成長し

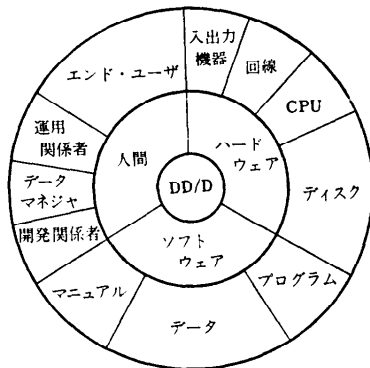


図-4 情報システムのイメージ

つつある。そのベースは、この 80 年代を通じて最も速いものの一つであるに違いない。ここで「情報システム」とは、ソフトウェアと同義でないことを強調しておきたい。それは、図-4 に示すように、ソフトウェア、ハードウェア、人間から成るものとする。DD/D はその中心に、Cincom 社の R. Weeks のいうように、情報システムの Brain として位置づけられる。

1983 年初めには ANSI X3H4 によって DD/D の標準化案がまとめられるという。DD/D は 80 年代のソフトウェアの最大の課題の 1 つであり、多くの成果が生れるものと期待される。わが国の研究者の精力的な研究を期待して筆をおく。

参 考 文 献

- (ANSI 75) ANSI/X 3/SPARC Study Group on Data Base Management Systems: Interim Report, ACM FDT 7, 2 (1975).
 (BRIT 77) The British Computer Society Data Dictionary Systems, Working Party Report, SIGMOD Report, Vol. 9, No. 4 (1979).
 (CHAN 81) Chan, P. and Shoshani, A.: Subject: A Directory Driven System for Organizing and Accessing Large Statistical Databases

- VLDB, Proc., pp. 553-563 (1981).
 (CURT 81 a) Curtice, R. M. and Jones, P. E. Jr.: Fundamentals of Data Element Definition, SIGMOD, Proc., pp. 49-55 (1981).
 (CURT 81 b) Curtice, R. M.: Data Dictionary: An Assessment of Current Practice and Problems, VLDB Proc., pp. 564-570 (1981).
 (DATA 78) DATAPRO: How to Evaluate and Select Data Element Dictionary/Directory Systems, EDP Solution, Datapro Research Corp. (1978).
 (DBSE 76) データベース専門委員会: データベース・システムに関する調査——データ・ディクショナリ/ディレクトリ——日本電子工業振興協会 (1976).
 (DIEB 79) Diebold, J.: Diebold Special Report. Infosystems, Vol. 26, No. 6, pp. 50-53 (1979).
 (GUID 74) GUIDE Dictionary/Directory Project: Requirements for the Data Dictionary/Directory within the GUIDE/SHARE Data Base Management System Concept (1974).
 (IKED 81) 池田秀人他: DBMS の DD/D の設計とその応用, 情報処理学会第 22 回全国大会予稿集 IL-4 (1981).
 (KENT 80) Kent, W.: Splitting the Conceptual Schema, VLDB Proc. pp. 10-14 (1980).
 (KENT 81) Kent, W.: Data Model Theory Meets a Practical Application, VLDB Proc., pp. 13-22 (1981).
 (LEFK 77) Lefkovits, H: Data Dictionary Systems, QED Information Sciences, Inc., Wellesley, MA (1977).
 (RISC 80) Risch, T.: Production Program Generation in a Flexible Data Dictionary Systems, VLDB proc., pp. 343-349 (1980).
 (TAMI 80) Tamir, M. 他: DBI A DBMS-Based Application Generator, VLDB Proc., pp. 335-342 (1980).
 (UHRO 73) Uhrowczik, P.P.: Data Dictionary/Directories, IBM Systems Journal, Vol. 12, No. 4, pp. 332-350 (1973).

(昭和 57 年 6 月 1 日受付)