

C/S情報系システムへのDOA適用

高橋まゆみ 南波則孝

日立システムエンジニアリング (株)

データ中心アプローチは広く実用化が進められてきたが分散データベースの構築や3階層C/Sアプリケーションへの適用技法が必要とされている。本論文ではこれらの新しい要件に対応するためのスパイラルなデータ分析手順と分析支援環境を提案する。特にデータ制約についてはデータ制約の種別を分類しドメインをベースにしたC/Sアプリケーションへの実装を行なった。最近1~2年間にいくつかのプロジェクトに新しいデータ分析環境を適用した結果、プロジェクトのデータ分析工数を削減することができた。また、データ分析の成果を活用してサーバのデータベース設計やクライアントのプログラム開発を標準化することができた。

Effective Data Oriented Approach For Client / Server Information Systems

Mayumi Takahashi Noritaka Nanba

Hitachi Systems Engineering, Ltd.

{takama,nanba}@iabs.hitachi.co.jp

DOA(Data Oriented Approach) has spread practically, recently new techniques are needed for developing distributed Data Base and three tiers C/S application. In this paper, we propose spiral data analysis process and environments for these requirements. Especially, we classified data constraints into some groups and implemented them using domain categories. As the result of applying effective data analysis environments to several projects in the last 1~2 years, We have reduced the man-power for data analysis and standardized Data Base design and program development for C/S information systems.

1. はじめに

データ中心アプローチ (DOA:Data Oriented Approach) は情報システムで扱うビジネスデータに着目した業務設計技法として広く実用化が進められて来た。

DOAは従来の業務プロセス中心の設計技法に対して、次の特徴を持ち、ソフトウェアの再利用、保守性の向上を狙っている。

- ・データモデルをベースにしたビジネスデータの一元管理
- ・データとデータに依存したプロセスのカプセル (部品)化

DOAの基本技法は既に確立されている[1]が実際のプロジェクトの適用にあたっては次の様な問題があった。

- ①現行データ調査に多大の工数がかかる
- ②データが分析されていく過程が難しい
- ③開発での効果が把握しにくい

これまでDOA適用効果としてホスト型システムでの開発事例は多く報告がある。[2][5] 現在、業務システムの構築はホスト型からC/Sシステムが中心となっており、分散データベースの構築や3階層アーキテクチャをモデルとしたC/Sシステムへの適用技法が必要とされている。

本研究ではこれらの新しい要件に対応するため効果的なデータ分析と適用法を検討した。第2章ではデータ分析手順を述べ、第3章でペーパレスのデータ分析環境を提示する。第4章ではC/Sシステムへの適用技法として特にデータ制約の実装について述べる。第5章では第2～4章で示したデータ分析を実施したプロジェクト事例評価を述べる。

2. DOA適用手順

日立ではシステム開発標準手順としてHIPACE (Hitachi high-PACE)がありその中でデータ分析標準手順を設定している (表1)。

表 1 HIPACEデータ分析標準手順

項番	工程	工程名	作業内容
1	DA1	現行業務データ調査	DB やファイルに含まれるデータ項目を調査して、名称、属性、意味などを統一的に調査する。
2	DA2	ドメイン分析	データ表現形式に着目し、データ項目定義の基になる値の集合を「ドメイン」として定義する。
3	DA3	業務データモデリング	データが表現する対象を「実体」として定義する。また実体間の関連を明示して「ER図」を作成する。
4	DA4	制約分析	データの表現形式、正当性の判断条件、導出ルール、用途 (目的) などを整理、検討する。
5	DA5	標準データ項目定義	「実体」と「ドメイン」を基に「標準データ項目」を定義する。
6	DA6	リポジット登録	標準データ項目、実体 (ER図)、ドメインをリポジットに登録するためのフォーマット検討、変換をする。

DA: Data Analysis、DB: Data Base、ER: Entity Relationship

従来はデータ分析手順DA (Data Analysis) 1から6までをフェーズドアプローチで行うことを標準としていたが、全体をボトムアップで進めると種々の問題が発生していた。

(問題)

- ①業務モデリングのデータ正規化に工数がかかる
- ②分析中判断できないデータがあると作業が止まる
- ③全体の整合性を保持するのが難しい
- ④調査票・定義票の記入が多い

それらの問題に対しデータ分析手順を2つのフェーズに大別しデータ分析を進める手順とした。(図1)

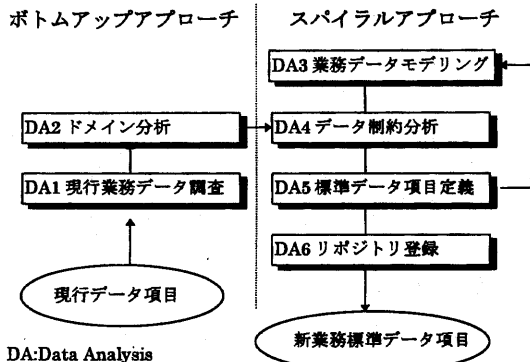


図 1 策定したデータ分析手順

- フェーズ 1 現行業務データ調査
～ドメイン分析
- フェーズ 2 業務データモデリング
～標準データ項目定義

ドメイン分析 (DA 2) はデータ項目の属性 (桁数・タイプ)、値域の分析を行うためデータ項目単位の分析となる。このフェーズは現行業務データを収集して既存データの問題分析を行う。さらに新業務の要件を加えることで、ビジネスをカバーする品質の良い標準ドメインを決定することができる。

一方、データモデリング (DA 3) でのエンティティ抽出は、モデル全体の整合性を取るの非常に難しいため、業務で必要とされる主要な管理対象を先行して決定し、徐々に標準項目を割り当て詳細化していく手順が効果的である。このスパイラルアプローチの過程でサブタイプの洗い出しやエンティティ間のリレーションを分析していくことができる。また、このデータモデリングのサイクルの中で新業務に必要なデータ項目を追加し、全体のエンティティとリレーションを新業務モデルとして作成していくことができる。

フェーズ 1 で定義されたドメインはデータ項目の設定時に標準属性 (桁数・タイプ) として利用でき、データ制約分析のベースとなる。

フェーズ 1 に引き続き、フェーズ 2 を現行データ項目を利用したボトムアップアプロー

チを進めるとデータの正規化作業に工数がかかるばかりでなく、現行業務仕様を反映するため、新要件を満たす業務のモデルを作成することはできない。この結果、新業務開発に利用できず、モデルを再作成する必要が生じる。

3. データ分析支援環境

2. で述べたスパイラルなデータ分析手順を実現するためデータ分析支援環境 (DAE: Data Analysis Environments) を開発した。本環境は Microsoft Access で実装され、データ分析手順 (DA 1～6) に対応した機能を持っている。データ分析後、完成したデータモデルは FIPS (米連邦情報処理規格) として認定されデータモデリング標準となっている IDEF1X (ICAM DEFinition 1 eXtended) [3] に準拠した米国製品 ERwin、日立開発支援製品 SEWB に連携することができる。

現行データは通常、既存のマスタファイル、データベースから直接、DAE にロードして分析を開始する。特に現行データ分析からドメイン分析は標準ドメインの抽出を効率よく分析するためのドメインの基本カテゴリを設定している。(表 2)

表 2 ドメインの基本カテゴリ

項番	カテゴリ	説明文
1	識別コード	管理対象を識別するための文字列 (値) であり必ず個々の値の指し示す対象が決められている。
2	識別番号	個々の値と管理対象の対応があらかじめ決められていない。おもに時々刻々発生する出来事を識別するために使用する。
3	区分	コードの一種であるが、ある基準による物事の分類を明らかにする目的を持つ。
4	フラグ	ある特定の条件を満たすか否かを示す。
5	日付	ある特定の時点の年月日を表わす。
6	時刻	ある特定の時点の時刻を表わす。
7	期間	ある特定の時点から次の時点までの期間を表わす。
8	時間	ある特定の時点から次の時点までの時間を表わす。
9	金額	円、ドルなどの通貨の額。
10	数量	個数、重量などを表わす数値。
11	名称	物事の名称を表わす名詞または短い句。
12	記述	名称より長い文字列であり、句あるいは文に相当する。

データ分析環境 (DAE) ではこのドメインカテゴリを利用して既存のデータを分類していく方式としている。(図2)

ID	項目名	カテゴリ	タイプ	制約	状態
133	郵便番号	郵便コード	X	14	0
134	郵便名簿	郵便番号	X	12	0
135	郵便店 T E L 市外	区分 フラグ	X	0	0
136	郵便店 T E L 市内	区分 フラグ	X	4	0
138	郵便区 CODE	郵便コード	X	2	0
139	郵便区 CODE	郵便コード	X	2	0
141	郵便区 CODE	郵便コード	X	2	0
144	郵便店 T E L	郵便番号	X	4	0
145	郵便店 T E L	郵便番号	X	4	0
146	郵便店区 CODE	郵便コード	X	4	0

図2 現行データ項目整理画面例 (一部)

データ項目をそれぞれのドメインカテゴリに大きく分類することで業務データの特性を整理することができ、カテゴリ別にドメイン分析が進められる。分析されたドメインは標準属性(桁数・タイプ)、利用目的、制約条件等が定義される。(図3)

ドメイン名: 368
ドメイン: TEL番号

カテゴリ: 区分フラグ
属性: 郵便・本属性: 0

用途(目的):
制約条件:

定義式:
文字数: 0
文字種: 0
重複: 0

図3 ドメイン定義票画面例 (一部)

また、データモデリングは主要なエンティティを先行して定義し、現行データから重複排除して分析対象となったデータ項目を順次割り当てていく方式としている。(図4)

エンティティのサブタイプの定義、関連するデータ項目の追加・変更が容易にできる。

サブタイプ区分: []
サブ(サブタイプ)属性番号: []
サブ(サブタイプ)属性名: []

識別キー: TEL 393 394 395 396 397 398 399 400 404 406
参照キー: TEL 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

図4 実体定義票画面例 (一部)

このデータ分析環境 (DAE) により、データ正規化の負荷を軽減しデータモデル全体をスパイラルに構築することができる。

エンティティ-データ項目-ドメインの関連はDAEで管理され、データモデルの整合性が保持できる。また、これまで理解しにくかったデータモデリングの過程をビジュアルに見ていくことができる。

4. データ制約分析とC/Sへの適用

データ分析を行った成果としてドメインとデータ項目は標準データ項目辞書として活用され、データモデルはデータベース構築の論理モデルとして開発に利用される。その応用技術は多く提案されている。[4]

しかし、データ分析で抽出したデータ制約の分類と活用についてはこれまで単一のデータ項目のバリデーション・ルールや編集条件の実装レベルに限られていた。[5][6]

本研究ではデータ制約をデータモデルに対応して分類し、C/Sシステムに対応した適用方式を設定した。

4. 1 データ制約の分類

データ制約の分類について標準規格は決められていないが分類例と分析手順は報告されている。[7]

本研究ではデータ制約はその表す意味と定義対象によって次の種別とした。(表3)

表3 データ制約の種別と定義対象

項番	分類	定義対象	制約の内容
1	形式制約	ドメイン・データ項目	データの型 (桁数・タイプ) を規定する
2	値制約	ドメイン・データ項目	データの取り得る値を規定する
3	導出制約	データ項目	導出先データが導出元データから算出される
4	関連制約	データ項目	複数のデータ間の関連によって規定される
5	存在制約	エンティティ	エンティティの属性となるデータの存在を保証する (Not Null)
6	参照制約	エンティティ	エンティティ間で参照キーの値が主キーの値として存在する

4. 2 データ制約の適用方式

4. 1 で分類したデータ制約が3階層C/Sアプリケーションモデルでどの階層に対して定義されるかを図6に示す。

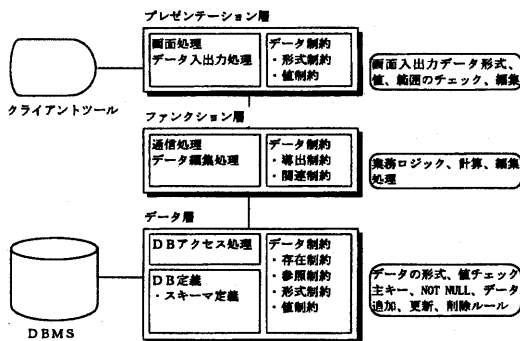


図5 3階層C/Sアプリケーションにおけるデータ制約の定義

(1) 形式制約・値制約

形式・値制約はドメインまたは単一のデータ項目によって決定される制約であり、プレゼンテーション層またはデータ層で定義される。

プレゼンテーション層ではデータ入出力時に入力データの形式、値域のチェック、データ層に引き渡すための編集条件として定義される。

データ層ではデータベースのカラム属性として定義される。

(2) 導出・関連制約

導出、及び関連制約は複数のデータ項目に関連する制約でビジネスルールとしてファンクション層に定義される。

(3) 存在制約・参照制約

データベースの追加・更新・削除時にリレーションを持つエンティティ間での整合性を保持する制約でデータ層に定義される。

上記(1)～(3)の適用法はデータ制約をどのC/Sアプリケーション層に定義すべきかを示したもので各層が物理的にどのマシン上にあるかには依存しないものとする。

5. データ分析事例の評価

これまでデータモデリングを実施してきたいくつかのプロジェクト事例を分析し新しいデータ分析手順の評価とデータ制約のC/Sアプリケーションへの実装例を述べる。

5. 1 データ分析の評価

過去1～2年の適用事例から次の4社の事例を分析する。

この中でS社、H社、O社は2、3で述べた手順とデータ分析環境を利用した。

A社は現行システムの画面帳票を収集しデータ正規化を行う手順をとった。

- ・ S社 製造業 (情報系システム)
- ・ H社 金融業 (情報系システム)
- ・ O社 製造業 (情報系システム
+ 人事システム)
- ・ A社 流通業 (情報系システム)

5. 1. 1 データ分析結果

はじめに各社のデータ分析対象となったりソース数とデータ分析の結果を表4に示す。

表4 データ分析対象リソース数

項番	対象内容	H社	O社	S社
1	データ項目数 (分析前 a)	5000	8350	3430
2	データ項目数 (分析後 b)	3800	3818	2033
3	エンティティ (d)	280	452	229
4	カテゴリ数 (e)	13	13	8
5	ドメイン数 (f)	489	1007	433
6	データ新旧比 (b/a)	0.76	0.46	0.59
7	エンティティあたりのデータ数 (b/d)	13.57	8.45	8.88
8	ドメインあたりのデータ項目 (b/f)	7.77	3.79	4.70

初期のデータ分析対象となったデータ数は3社で差異がある。S社は過去に整備した標準辞書があったため項目数が少なくなっているが、分析対象のデータ数は4社の対象業務の範囲によって異なっている。

データ分析後、標準化されたデータ項目は2000から3000項目となった。また、1エンティティあたりのデータ数は平均10項目程度となった。1ドメインに対応するデータ項目数は5項目程度となった。

2. で示したドメインカテゴリ別のドメイン数を表5に示す。

表5 カテゴリごとのドメイン数

カテゴリ	H社	カテゴリ	O社	カテゴリ	S社
識別コード	53	識別コード	85	識別コード	141
識別番号	45	識別番号	59	識別番号	14
区分	267	区分	674	区分	224
フラグ	44	フラグ	60	—(*1)	—
日付	9	日付	7	日付	23
時刻	9	時刻	3	時刻	9
期間	5	期間	7	期間	1
時間	3	時間	5	—(*2)	—
金額	6	金額	7	数値	9
数量	21	数量	45	—	—
名称	7	名称	16	文字	12
記述	7	記述	11	—(*3)	—
その他	13	その他	28	—(*4)	—
	489		1007		433

*1: フラグは区分としてまとめた

*2: 時間は期間としてまとめた

*3: 名称、記述は文字としてまとめた

*4: その他は作らない

カテゴリの設定については各社で粒度が異なりS社ではフラグ→区分、時間→期間、金額・数量→数値、名称・記述→文字に統合した。これはアプリケーションシステムでのドメインの活用レベルに依存するもので細分化したカテゴリでの識別効果（例 単位の付与、チェックルール、データ変換ルール）があるかによって決められる。

カテゴリ別のドメイン数は3社ともにコード系（識別コード、識別番号、区分）は全体の約8割を占め、その中でも区分に関するドメインが5から6割を占める結果となった。

5. 1. 2 データ分析工数の分析

次にデータ分析の作業時間の比較を（図6）に示す。

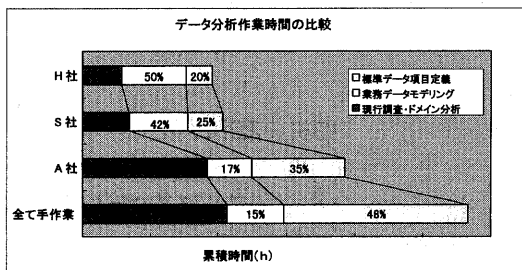


図6 データ分析作業時間の比較

この中でA社は現行システムの画面・帳票を収集し、正規化を行う手順をとったプロジェクトである。（データ分析対象数：約3000項目、分析後のデータ数：約2000項目）A社のケースでは調査票は電子化されて管理されており、データモデリングツールを利用しているため、データエントリやモデル図作成の工数はS社、H社とほぼ同等と考えられる。

この比較から2.で示した新手順を利用したS社及びH社はA社のケースに対して約50%の工数でデータモデリングを行うことができた。

A社の事例をベースにすべて手作業で行った場合を想定すると、新手順とデータ分析環

境の効果から全体工数は約1/3になる。
さらにデータ分析の手順と分析環境の効果をフェーズ別に分析するため、各工程の工数の割合を(図7)に示す。

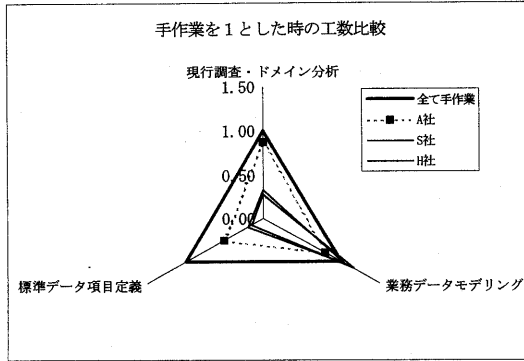


図7 手作業を1とした時の工数比較

新手順と分析環境を利用したS社とH社で特に現行調査・ドメイン分析、標準項目の定義が効果が大きいことが分かる。データモデリングに関してはほぼ同等の工数となり、分析支援環境だけでは十分な効果が得られていない。これはデータモデルの詳細化と品質の確保にはデータモデリングの基礎知識の他、業務知識を有すること、そのため洗練された正しいモデルを作成するために多くのレビュー時間を必要とすることが要因と考える。

これらのデータモデリングの効率化には分析者の思考を助ける業務知識の支援、データモデリングの不良摘出を支援する機能などが今後の課題である。

5.2 データ制約の実装評価

4. で述べたデータ制約のC/Sアプリケーションの実装についてS社の事例を示す。

S社では情報系システムとして従来はホストオンラインシステムであった売上情報システムのダウンサイジングを進めた。システム構成はサーバにDBMSを持ち、クライアントに簡易型GUI開発ツールを搭載してOLTP (Online Transaction Processing) を行う業務である。

S社では5.1.1 (表5) に示したドメイ

ンをベースにデータ制約を分析しサーバ、及びクライアントに実装した。今回の実装にあたってドメインカテゴリ別にデータ制約の設定ルールを決定した。(表6)

表6 ドメインカテゴリ別データ制約の設定表

設定内容	ドメイン								データ項目								備考
	コード系		時間系		文字		数値		コード系		時間系		文字		数値		
カテゴリ名	1 職別 コード	2 職別 番号	3 区分	4 日付	5 時間	6 期間	7 文字	8 数値	1 職別 コード	2 職別 番号	3 区分	4 日付	5 時間	6 期間	7 文字	8 数値	
サーバ	バリ デー ション ル ール 有 効	カ ラ ム 属 性	ラ ム 属 性	○	○	○	○	○	○	○	○	○	○	○	○	○	形式のみのチェック
	初期値			○	○	○	○	○	○	○	○	○	○	○	○	○	NULL、スペースなど
クライアント	バリ デー ション ル ール 有 効	カ ラ ム 属 性	ラ ム 属 性	○	○	○	○	○	○	○	○	○	○	○	○	○	→
	初期値			○	○	○	○	○	○	○	○	○	○	○	○	○	→
	表示書式								○								→有効桁ゼロサプレス
	編集書式			○	○	○	○	○									ハイフンの追加など
	ラベル、ヘッダ								○	○	○	○	○	○	○	○	

○：制約設定箇所、「空白」：ドメインの設定を継承
→：ドメインの設定を継承、カラムに着く制約は、リプレースする

(1) バリデーションルールの実装

カラム属性はサーバのDBMSのドメインの制約として実装した。

クライアントではドメインカテゴリ別に次の設定方式でバリデーションルールを実装した。

- ①コード系はすべてドメインの制約を実装
- ②時間系、文字・数値系はドメインの制約を継承するがデータ項目の制約でリプレース可能

これらはドメインカテゴリの特性に応じてデータ制約の実装方式をさらに詳細に検討したものでドメインを有効に活用することができた。

(2) 初期値

サーバのDBMSにはカラム初期値としてドメイン単位の設定を行なった。クライアント側では、コード系はドメインを必ず継承するとした。

(3) 表示・編集書式

クライアントの表示書式（表示上のゼロサプレス、有効桁数埋め、金額桁区切りのカンマなど）は、'数値'カテゴリーのみに設定した。

(4) データ関連制約・導出制約

S社の売上統計に利用する導出ルールは種別が非常に多く複雑であった。また、製品種別間の関連制約や製品販売ルールと顧客との関連制約はすべて分析をすることはできなかったがファンクション層での実装は再利用性の高いデータ制約をサブプログラムとして部品化した。サブプログラムにはデータベースの参照を含んでいる。関連制約の中でも特にエンティティ間のデータ項目に関する制約は分析のレベルやプログラムでの実行制御をどの様に実装するか、まだ課題が残されている。

S社ではデータ分析の結果として得られたデータモデル、標準データ項目、ドメイン、データ制約を活用してデータベース設計及びプログラム開発を行なった。プロジェクト全般にわたりデータ管理や開発の標準化を図ることができた。

6. まとめ

本研究では効果的なデータ分析手順とデータ分析環境を提示し、実際のプロジェクトでの適用成果をまとめることができた。

また、データ分析から抽出されるデータ制約についてC/Sへの適用法を示し、アプリケーションの事例を示した。

今後はさらに実プロジェクトへの適用を広げるとともに次の課題に取り組みたい。

- ①データモデリングの品質確保の方法
- ②データモデリングの分析者の思考を助ける業務知識の提供法
- ③特に複雑なデータ関連制約及び導出ルールの実装
- ④分散オブジェクト開発でのデータモデリングの適用法

7. 謝辞

本研究をまとめるにあたり御協力いただきました日立製作所ビジネスシステム開発センタの方々へ感謝致します。

8. 参考文献

- [1] 堀内一：「データ中心システム設計」（1989、オーム社）
- [2] 株式会社データ総研 編：「データ中心システム設計の実際」（1995、オーム社）
- [3] 松本聡：「IDEFIXリレーショナル・データモデルの新しい表現法」（1996、日経BP社）
- [4] 斎藤直樹：「データモデル設計とRDBMSへの実装」（1995、リックテレコム）
- [5] 森岡洋介 他：「データ中心アプローチを応用した標準データ項目辞書の開発とその活用方法」（1993.11、日立評論VOL.75）
- [6] 大野治他：「データ中心型ソフトウェア開発技法」（1994、電学論C、114巻6号）
- [7] IRM研究会編：「データ中心システム分析と設計」（1996、オーム社）
- [8] 堀内一：「情報資源辞書システムIRDSの標準化」（1996.7、情報処理学会VOL37NO.7通巻377号）
- [9] 徳鷹良介：「データベース関連技術の標準化の概要」（1996.7、情報処理学会VOL37NO.7通巻377号）

Microsoft Accessは米Microsoft Corporationの商標です。

ERwinは米国LogicWorks, Inc.の登録商標です。

SEWBは株式会社 日立製作所の製品名です。