

# 既存システム分析による情報整備手法

町原 宏毅\*、川下 満\*、丸山 則夫\*\*  
(\*NTT情報通信網研究所、\*\*NTT情報システム本部)

情報システムは、開発の優先順位や規模の問題により、業務の領域や目的に応じて別システムとして段階的に開発されている場合が多い。そのため本来同じであるべき情報が異なる形式で表現されているものもあり、必要に応じてデータ変換を行い、それらを流通させながら、企業全体の活動を支援するような情報システムの構造となってしまう。

そこで、本稿では、データ相互の関係が複雑な複数の情報システムを分析し、計画的に全社的に統一した情報システムを構築していくための手順について記述する。

## An Information Administration Methodology by Analyzing Existential Systems

Hiroki MACHIHARA\*、Mitsuru KAWASHIMO\*、Norio MARUYAMA\*\*  
\*Network Information Systems Labs.,NTT、\*\*Information Systems Headquarters,NTT

1-2356, Take, Yokosuka-shi, Kanagawa, 238-03 Japan

There are many resemble information systems that are developed on every application scope or purpose because of the problems of development priority or scale. On these information systems, because same informations are represented different styles, enterprise activities are operated by exchanging and distributing data between information systems.

So in this paper, we describe a methodology for making intentionally integrated information systems that have complex relations between information systems by analyzing existential systems.

## 1. はじめに

情報システムの普及にともない、年々その数は増加の一途をたどってきている。さらに、開発の優先順位や規模の問題やセキュリティなどの問題により必ずしも統合された情報システムとして構築されず、業務の領域や目的に応じて別システムとして段階的に開発されてきている。そのため、お互いが共通の情報を持ち、必要に応じて流通しあいながら、企業全体の活動を支援するような情報システムの構造となっている。

さらに近年ダウンサイジングによりエンドユーザのコンピューティングが盛んに行われるようになり、誰もがデータを検索、加工できるような機構が広まりつつある。このような状況の中で、なんら標準的な規約や制約もなく、開発を進めていくことは、将来のデータ流通の機構が複雑になり、流通できないような状態になってしまうことも起こりうる。さらには、データベース化されている実際の値の誤謬が生じ、企業としても信頼性が失われてしまうことになりかねない。

そこで、本稿では、そのデータ相互の関係が複雑な複数の情報システムに対して、既存の情報システムを分析し、それらの統合的なデータモデルを作成し、それに基づいて情報システムを順次構築/更改していくことにより、全社的に統一のとれた情報システムを構築していくための手順について記述する。

## 2. 従来技術と問題点

複数の情報システムにより企業全体の活動を支援するような構造に合、その中で独立した（共通の情報を持たない）情報システムはほとんど存在せず、何等かの情報システムと関連を持って運用されている[1]。そのため図1の例に示すように情報を認識する単位である1つの実体（ここでは「回線」）がさまざまな業務（受注業務、料金業務、設備業務、etc）で利用されている。その結果、それらの業務を情報システム化した場合、本来同じ実体が異なる捉え方をされているケースが存在し、その実体を論理/物理データベースに展開した場合、図2に示すように本来同じものを表現しているデータ項目が異なる表現形式（名称、データ型、桁数、etc）で定義されてしまう事になる。そのためユーザ要件の変化により、情報システムに共通的に保有されているデータ項目

を変更させる必要が生じた場合、どの情報システムのどこを直せばよいかを特定できなかったり、データ流通時のデータ変換に負荷がかかってしまったりする。

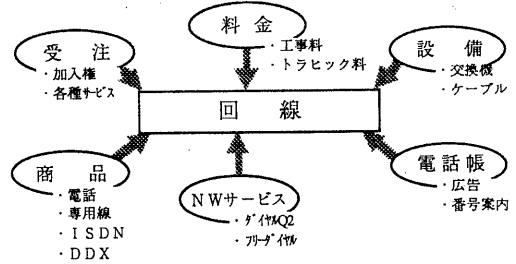


図1 実体「回線」に対する様々な捉え方

具体的には、以下のような状況となっている場合がある。

- ・同じ意味のデータ項目名称が異なる。
- ・同じ意味のデータ項目の属性、桁数が違う。
- ・同じデータ項目を必要以上に重複保有する。

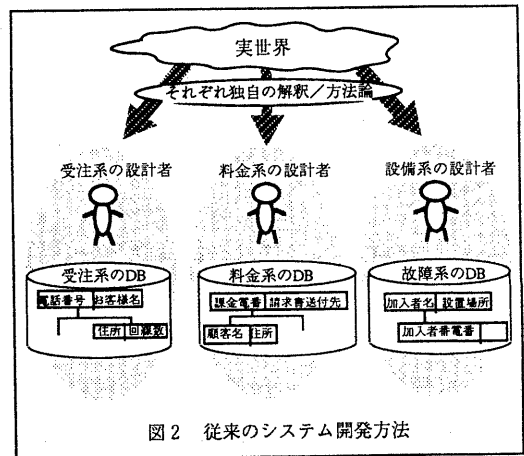


図2 従来のシステム開発方法

さらに今後の情報システム開発の動向としては、各部門が独自にカスタマイズできるように、また各部門が必要に応じて複数システムをまたがる情報を収集しやすくする必要があります。

そのようなことを実現するために、情報システムを持つデータベースの定義情報を標準化しようとする動き[2]と、データベースの定義情報はいじらずにデータ流通時のゲートウエーで変換をかけようとする動き[3]がある。ここでは前者のデータベースの定義情報を標準化していくための手法について述べるが、この中にも、従来はトップダウンで作成された統一的なモデル（企業モデルと呼ばれる）からその

一部を切りだし、そのデータモデルから情報システムのデータベースを構築していく方法[4]や、リエンジニアリングとして、既存の情報システムを一度概念モデルで表現し直し、新たな情報システムを構築していく方法[5]などが考えられている。

しかし、前者の場合は、新たな情報システムを開発したり、全く作り直したりする場合には良いが、既存の情報資産がうまく活用されず、アプリケーションプログラム自体の大幅な見直しに伴うため、実際には多くの経費がかかってしまう。また後者の場合では、複雑に関係しあった複数の情報システムにおいてはうまく適応できないという問題があった。

### 3. 本手法の概要

前節で挙げたような複雑に関係しあった複数の情報システムにおいて、個別の分析により作成されたデータモデルを統合する手順を示すことは非常に困難であった。そこで本稿では、さまざまな表現をしている情報（実体や属性）に対して、標準的な名称を付与し、それらを構成する用語の構成を元に行うことにより、具体的な統合データモデルの作成手順について示す。さらに一度に全ての情報システムを標準化してしまおうというのではなく、長期的な計画の中で優先順位を付け、序々に情報システムを標準化していくための指針を示す。

情報整備の手法としては、図3に示すように「システム計画」、「情報整理」、「データ分析」、「データ標準化」、「適用」、の5つの工程に分けて整理していく。

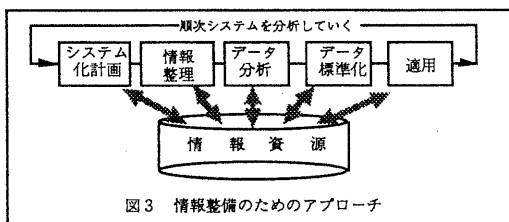


図3 情報整備のためのアプローチ

各工程の概略の作業を以下に示す。

#### ・システム化計画：

分析対象システムを全社的観点から捉え、長期的標準化の計画と、具体的データ分析に関する全体の作業線表を作成する。

#### ・情報整理：

分析対象システムが保有するデータを洗いだし、そのデータ項目の持つ情報を整理する。

#### ・データ分析：

データ項目の持つ従属関係をもとにデータを整理し、正規化する。

#### ・データ標準化：

自システムのみならず、他システムとの関係で統一しなければならない情報を明確化し、その対応を管理する。

#### ・適用：

データ分析作業に関する成果物をディクショナリに格納し、具体的な実システムへの適応方法を検討する。

### 3.1 システム化計画

「システム化計画工程」では、まず、全社的な情報システムの構成を調査し、どの情報システムがどの企業活動をサポートしているのかを明確化する必要がある。これによりお互いに共通の情報を保有している可能性のある情報システムがグルーピングされることになる。次にそれらの情報処理システムに対して、どのように事業展開を進めていくかについて明確な方針をたてる。

具体的には、分析対象としての情報システムだけでなく、その情報システムと関連のある領域全体を見渡し、その領域に対する業務やデータを長期的に見てどのようにしていくべきなのかといった経営的判断が必要となり、以下の事項を明確化しておく。

#### ・該当情報システムの問題点の抽出

#### ・標準化の目標/目的

#### ・該当情報システムを標準化させることの効果予測

また、この段階で実際に分析を行うための具体的線表/手順を定める。

### 3.2 情報整理

「情報整理工程」では、分析対象システムが保有するデータ項目を洗いだし、そのデータ項目の持つ情報を整理する。また、情報システムの更改時に、新規に追加されたり、変更/削除されたりするユーザ要件が存在する場合は、そこで扱われるデータ項目についても整理する。

現在の情報システムを見てみると、設計情報やその定義などを記述したドキュメントが整理されていないケースが多々存在する。そこで、主な作業として、それら設計情報を洗いだし、定義を明確化する。

この時、実際にデータベースの定義情報は、さまざまな制約（性能、メモリ容量、格納容量、等）を満たすため、データを冗長に保持していたり、定義名称などもテーブル（あるいはセグメントやレコード）毎に一貫していない場合が多い。

そのためこの工程では、具体的には以下のような事を整理しておく必要がある。

- ・データ項目の意味を明確化する。
- ・データ項目の名称を標準化する。
- ・同音異義の項目を整理する。
- ・異音同義の項目を整理する。

データ項目の名称を標準化するために、Durellの命名規則[6]を基本に、日本語表現を拡張し、計算機化することにより分析作業を支援するツールを開発した[7]。

Durellの命名規則とは、図4に示すようにデータ項目を構成する単語（ここでは用語と表現する）を役割別に3つのタイプに分類し、データ項目命名規則を定める。ここではデータを標準化するために、以下の要件を満たすことを定義している。

- ・業務やDBに精通していないエンドユーザにもわかりやすいこと。
- ・同じデータ項目には同じ名称を付与する。
- ・異なるデータ項目には異なる名称を付与する。

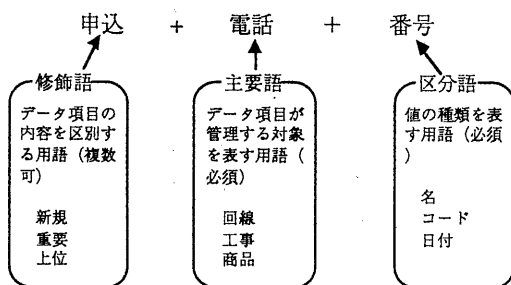


図4 データ項目の命名付与規則

このようにまず、既存の情報システムで表現されているデータ項目の名称を標準化し、用語単位にデータ項目を区切ることにより、データ項目の類似性を判断しデータ項目の同音異義や異音同義について整理することにする。

### 3.3 データ分析

「データ分析工程」では、前述の「情報整理工程」で明確になった分析対象のデータの冗長性をなくす

ために、データ項目の持つ従属関係をもとにデータを整理し、正規化作業を実施する[8]。既存の情報システムでは、以下のような構造でデータを定義している場合がある。

- ・1つのデータグループ（テーブルやセグメント、レコード、等のデータ項目を集約する単位）に、複数の実体が含まれている。
- ・同じデータ項目が、複数のデータグループに含まれている。

ここでは、分析対象のデータに対して冗長性を排除し、成果物として分析対象としている情報システムのデータモデル（実体関連図）を作成する。

具体的には、以下のような順序で行う。

(1)「情報整理工程」で整理されたデータ項目の一覧表を用意する。

このとき、データ項目一覧には、データグループ名、データ項目名、標準データ項目名、意味定義の順で表現されているものとする。

(2)類似のデータ項目をグルーピングする。

類似なデータ項目を検出するために、データ項目を構成する用語に対し、あらかじめその用語の持つ表記法の違いや類似の内容を持つものをグループ化して、図5に示すように階層的に表現しておく。

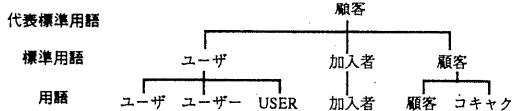


図5 デイクシオナリ内の用語の構成

このような用語の構造のもとで、主要語が類似しているデータ項目を分類すると、（ユーザ名/ユーザ氏名/顧客住所/加入者名称）のような論理モデル上のデータ項目は類似のものとして判断され、抽出されるため、その中でデータ項目の類似性を検証することにより、全体の分析作業の効率を上げることができる。また全ての作業に対して共通の用語の構造をもったデイクシオナリを使用することにより、複数の作業が行った場合でもその結果を揃えることができる。

(3)同じグループ化された項目の集合体に名称を付与し、それを実体として定義する。

ここでは、主要語のシソーラスの代表的な用語をそのまま実体名として付与する。

(4)同じグループのデータ項目についての同一性を確認し、実体を構成する属性として定義する。

類似のデータ項目として抽出された項目について同音異義や異音同義を検証し、同音異義であれば、意味説明に合った標準データ項目名に変更する。また、異音同義であれば、それらの標準データ項目名を一致させる。

(異なる実体の中に本来同じデータ項目が存在するケースや、同じ実体の中に本来異なる実体に含まれるデータ項目が存在するケースも有り得るが、それらについては、データ項目の分類作業を繰り返し行い、さらに以下の正規化の作業の中で明らかにしていく)

(5)同一実体内に含まれる属性に対して、それらの間の従属関係を元に、正規化を行う。

-正規化の具体的手順について論文で提案されており、ここでは省略する。

(6)分析結果を整理し、分析対象システムのデータモデル(実体関連図、実体定義票、実体属性票)としてまとめる。

### 3.4 データ標準化

「データ標準化工程」では、その時点の統合データモデル(すでに同じ領域の統合データモデルが作成されていない場合は、その分析対象システムのデータモデル自身が統合データモデルとして位置付けられる)と分析対象システムのデータモデルとの関係を明確化し、分析対象システムを含めた統合データモデルを作成する。

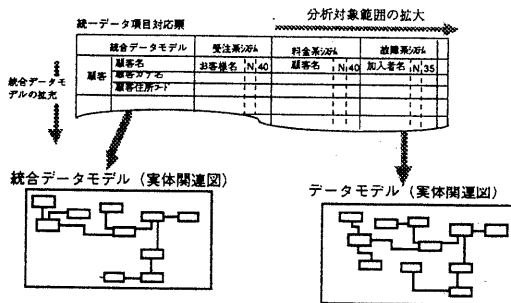


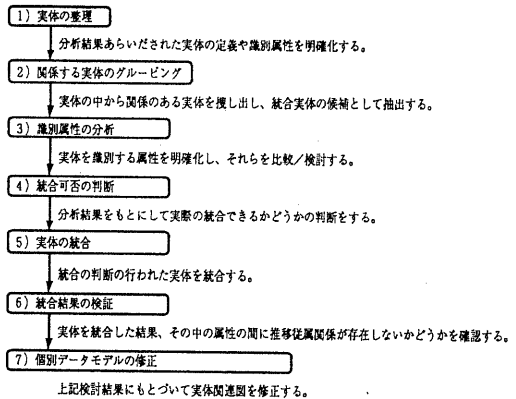
図6 データ標準化の概要

ここで、統合データモデルとは、図6に示すように企業が扱う情報に対して標準的なものとして定義したものであり、個別システム間で重複してもたれる情報を横通しするための参照モデルとして位置付けられる。つまり、新たなシステムの場合はそれをもとに構築されることを求めるが、既存システムの

場合は、その間の重複度を管理するためのインデックスのように扱われる。情報整備が進み、標準化が行き渡った形態では、それら横通ししたデータ項目全てが統一されることになる。また、統合データモデルは、既存システムの分析を行いながら拡充していくため、分析を行いながら必要に応じて追加/修正が加えられる。

#### 3.4.1 統合データモデルの作成手順

複数のデータモデルを統合する方法としては、いくつか示されているが[1][9]、ここでは、実体名を1つの用語と見なし、そのシソーラスの構造から統合候補を抽出し、統合を進めていく手順について述べる。複数のデータモデルを統合する場合、同じ実体かどうかの識別や同じ属性かどうかの識別を行うのは、さまざまな判断が必要となり、一元的に行うことは非常に難しくなる。そこでここでは、個別に作成された独自のデータモデルの実体名や属性名で扱われている用語をもとに分類し、1つの共通的な統合データモデルを作成する手法について述べる。



##### 1) 実体の整理

分析結果あらいだされた実体について以下の事項を整理する。

・実体名；

実体を表現する名称について明確な用語で表現する。シソーラスを表現した辞書の中に定義されていない用語を使用している場合は、共通の辞書の管理者がシソーラスを定義する。

・実体定義；

実体の表現する内容について、とりうる値の範囲や生成するタイミング、について明確化する。

- ・実体の識別属性；  
どのような属性によって識別されるかを明確化する。
- ・実体間の関連；  
他の実体とどのような関連があるかを明確化する。

## 2) 関係する実体のグルーピング

実体の中から「関係のある実体」を捜し出し、統合実体の候補として抽出する。関係のある実体の見方については、以下のように行う。

- ・実体名が類似のもの  
実体を構成する用語のシソーラスを考えた場合に同一の構造内に存在するもの。
- ・類似の業務活動の中で利用されうる実体  
別々に整理された実体であっても同じ業務では同じ様な実体を扱うことになる。

## 3) 識別属性の分析

実体を識別する属性を明確化し、その識別属性の持つ形式（名称、型、桁数、属性説明、etc）を比較／検討する。

- ・属性名；属性を表現する名称について明確な用語で表現する。
- ・型；属性を表現するデータの型（全角、半角、数字、英数字、etc）を明確にする。
- ・桁数；属性を表現するのにどれくらいの桁数が必要としているのかを明確化する。
- ・属性説明；属性の表現する内容として、実際にとりうる値の範囲やデータの生成するタイミング、etc について明確化する。

## 4) 統合可能の判断

実体の統合については、実体の持つ性質などによって統合の判断を行っているものもあるが、実体の認識自体が分析者により異なるため作業による結果の食い違いが生じてしまう。そこで、ここでは、「関係する実体のグルーピング」の結果同じ実体と見なせるもののうち、「識別属性の分析」で同じ識別属性を持つもの同士を基本的に統合させることにする。

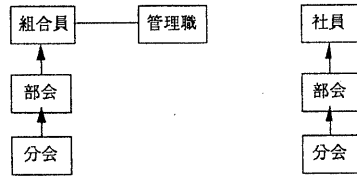
但し、以下のような場合は、統合させない。

- 1) 一方の実体にもみ関連があるような実体  
-このような実体を統合させると、統合させる前の実体と関連のある実体との区別ができなくなる。

例えば、実体「組合員」と「管理職」がそれぞれ図7（a）に示すような属性を持ち、（b）に示すように「組合員」にのみ他実体と関連があるような場合、（c）のように統合して表現してしまうと、全ての「社員」がいずれかの「部会」に属するような表現と区別できなくなってしまう。

組合員				管理職			
氏名	社員名	社員住所	自宅電話番号	氏名	社員名	社員住所	自宅電話番号
							管理評価

(a) 実体「組合員」と「管理職」の属性



(b) 実体関連図

(c) 統合させた場合の実体関連図

図7 異なる関連を持つ実体のどうしの統合例

## 4) 自己属性に共通項目のない実体

-統合実体中の属性として値の埋まる属性と埋まらない属性とに分かれてしまう。

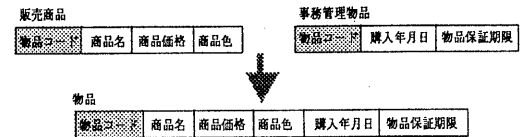


図8 共通項目のない実体どうしの統合例

## 5) 実体統合

- 統合された実体の名称、実体説明を定義し直す。
- 統合された実体に含まれる属性を見直す。
  - ・同じ属性は、同じ形式（名称、型、桁数、説明）に統一する。
  - ・異なる属性は、異なる形式（名称、型、桁数、説明）にする。

### ・実体名

統合する実体を総称する名称を付与する。

### ・実体説明

統合した実体を説明し、統合の観点や統合前の実体に関する情報も付与する。

### ・属性の区別

同じ実体として統合された属性の中で同じ項目が存在しないかどうかを確認して、定義する。ある実体では1つの属性を統合対象の実体では、複数の属性で表現していたりすることがある。

### 6) 統合結果の検証

実体を統合した結果、その中の属性の間に推移従属関係が存在しないかどうかを確認する。

実体を統合させることにより、お互いの属性の間に推移従属関係が生じる場合がある。このような状況になった場合は、もういちど統合された実体を分解して第3正規形にする必要がある。

例えば、

実体1 {A+B+C+D} と実体2 {A+F+G}  
(Aが識別属性を表す)

を統合させた場合、属性「C」と属性「G」の間に推移従属関係が生じる場合、新たに実体3を作成し、第3正規形にする。

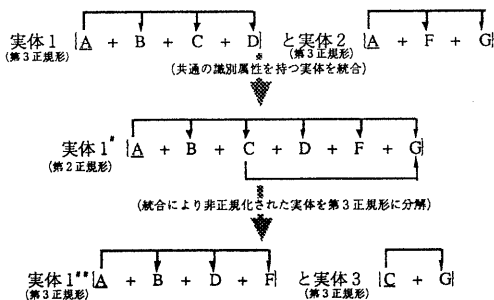


図9 遷移従属のある実体の分解例

### 7) 個別データモデルの修正

分析対象となった情報システムのデータモデルをもとに統合データモデルを作成した後、統合データモデルとの違いをもとに、対象データモデル自身の見直しを行う。

## 3.5 適用

### 3.5.1 ディクショナリへの登録

「適用工程」では、データ分析作業に関する成果物をまとめ、情報資源管理システムのディクショナリに格納する。このディクショナリでは以下の要件を満たす必要がある。

- ・情報システムの論理/物理DB情報の格納
- ・情報システムのデータモデルの管理
- ・情報システムの論理/物理DB情報とそのデータモデルとの対応管理
- ・統合データモデルの管理

- ・情報システムのデータモデルと統合データモデルとの対応管理

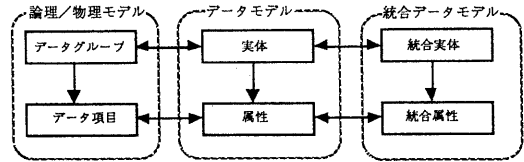


図10 ディクショナリの内部構造

上記情報について、お互いの情報から連鎖的に情報を参照できるようにしておく必要がある。これにより、具体的に情報整備を実現するために現状どの程度食い違いがあり、何をどう合わせていけば良いのかを明示することができる。

### 3.5.2 実システムへの反映

既存の情報システムを分析し、統合データモデルを作成したとしても、それがすぐ既存のものに反映されるものでもなく、統合データモデルに合わせるための稼働や効果によって適応の度合は異なる。

上記手順の分析により、ディクショナリには、情報システムの論理モデルとデータモデルとの対応をとり、さらにデータモデルと統合データモデルとの対応が管理されていることになる。そこでここでは、既存の論理モデルを統合データモデルの表現に合わせる場合にどの部分を修正しなければならないかが明らかになる。そこで、以下の事項について定量的に示し、その全体のバランスにより情報システムの更改を行っていく。

#### (1) 論理モデルで表現されているデータ項目数

-本分析対象となった項目数

#### (2) 論理モデルと統合データモデルの間でデータ項目の表現形式(名称、型、桁数)の食い違っている項目数

#### (3) 統合データモデルに合わせるための稼働

-該当データ項目をどのプログラムがアクセスしているかを導出し、全体の修正量を算出する。

#### (4) 統合データモデルと食い違うデータ項目のうち統一させるデータ項目数

-統合データモデルに合わせる稼働とその効果を定量的に見積り、経営的判断で決定する必要がある。

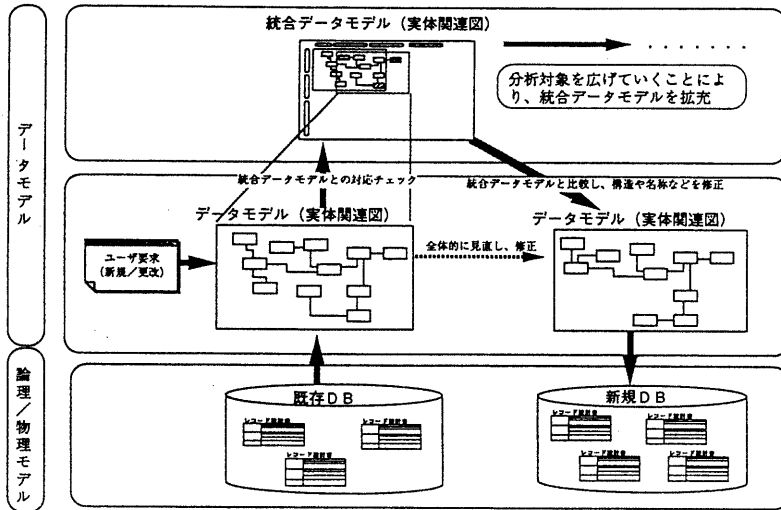


図 1 1 情報整備の流れ

このようにして、既存の情報システムのデータと統合データモデルとの差異について定量的に示し、統合データモデルに合わせる稼働と効果を比較しながら情報システムに対して適応していく。以上述べてきた情報整備の全体の流れを図 1 1 に示す。

#### 4. まとめと今後の課題

既存システムを分析した結果を基にして統一的なモデルを作成し、それを今後の情報システム全体の整備にうまく活用していく手法について述べた。この手法を用いることにより、序々にはあるが、作成される情報システム自体が標準化され、情報システム間の情報流通がスムーズに行われるようになる。

今後はデータ標準化の程度を定量的に表現する指針を示し、情報システムの新規開発/更改についての経営的判断を支援できるような指針を示せるようにしたい。さらに今回はデータについてのみ述べたが本来データ整備に置ける問題点は、業務中心の設計を行っていた事による場合が多く、その部分を根本的に見直さなければならない。そこで今後は業務とデータとを含めて体系的に表現できるような手順に拡張していきたい。

#### 参考文献

- [1]鈴木：情報基盤のためのデータベース再構築の7つのアプローチについて、情報処理学会データベースシステム研究会，p.223～232,,1993
- [2]関根、川下、町原、中川：体系的なDB構築のための用語辞書を用いたデータ標準化手法、情報処理学会論文誌，p.457～467,,1993
- [3]池田、石垣、村田：DB流通の基本方式について、情報処理学会第46回全国大会，p.4-97～4-98,,1993
- [4]J.Martin：Information Engineering，Book：Introduction,Prentice-Hall,1989
- [5]味村、山田、堀内：データベースシステムの設計と開発、オーム社、1983
- [6]Durell,R.W.：データ資源管理、日経マグロウヒル,,1987
- [7]町原、黒川、関根：データ標準化ツール(DBprompt/NAME)を用いたデータ分析手法、情報処理学会第44回全国大会，p.4-235～4-236,,1992
- [8]P.P.Chen:The Entity-Relationship Model;Toward a Unified View of Data，ACM TODS，Vol.1，No.1，March，1978
- [9]F.A.Bernstein:Synthesizing Third Normal Form Relation from Functional Dependencies，ACM Transactions on Database Systems，Vol.1,No.4,,1976