

4K-1

# 上流工程におけるエンティティ分析の留意点

林 恵美子  
富士通株式会社

森田 功 橋本 恵二  
システム生産技術センター

## 1. はじめに

最近、システム開発の上流工程でデータ分析を実施する際、データモデル(エンティティ-リレーションシップダイアグラム)を用いた方法が広く使われている。我々もその考え方も採用した独自の分析技法を開発し、実践している。

実際にデータ分析を進めることは容易ではない。それは、同じ業務についてのデータモデルでも作成する人によって異なる、というところに現れている。この原因として、エンティティ自体の認識が難しいということが考えられる。エンティティの決定は、データ分析の目的や環境、さらには分析範囲などによっても影響される。

我々は、いくつかの企業におけるデータ分析の実践経験から、このエンティティの問題についての整理を始めた。本文ではその試みを簡単に紹介する。

## 2. C-NAP IIのデータモデル

我々の使用しているデータモデルは以下のようなものである。

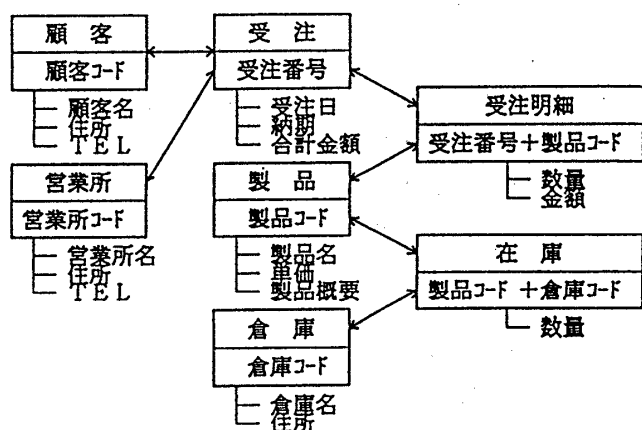


図1 C-NAP IIの概念データモデル

システム要求分析技法C-NAP IIで規定する概念データモデル(図1)を使用する。C-NAP IIは、利用者の要求を分析するための技法で、データ分析の他に、問題分析、業務機能分析などの技法を含む。この概念データモデルは以下のような特徴をもつ。

- ①エンティティをボックスで表現する
- ②・ボックス間の矢印でエンティティ間の二項関連(リレーション)を表現する
  - ・エンティティのサブクラス概念(is\_a 階層)も表現できる

## 3. データ分析の問題点

データ分析を行う上で、その企業の中にあるエンティティを引き出してモデル化することは重要であり、かつ困難な作業である。特に、エンティティを決定する上で判断に迷うことが多い。

この原因として、企業のエンティティの多様性が考えられる。企業が認識するエンティティはその企業によって異なり、レディメイドに業種や業務で同一であることはない。例えば、「受注」などと言った場合、販売管理業務ならどの企業も同じものを指しているようである。しかし、ある企業では注文全てについてを受注というが、他の企業では、引当ができたもののみを受注といい、在庫がない場合の注文の扱いが違うというようなことが頻繁にある。

このような業務特性の他に、その企業の他のシステムからの制約、システムに対する要求など、様々な要因から企業のエンティティは多様性を持つ。このため、分析の専門家が勝手にその企業のエンティティを決め込むことはできない。実務担当者の分析への参加があって初めて、その企業のエンティティを引き出すことができる。正規化を

をすれば、データモデルができるというような安易なものではない。

#### 4. 問題への対処

これらのエンティティについての問題に対して、我々はいくつかの実験経験から、問題の整理の試みを始めた。問題には同様のアプローチで対処できる、いくつかパターンがあると考えられる。それらの事例について、それをパターン毎に整理し、そのときの判断の基準や対処の方法がまとめられるだろう。

最終的には、エンティティの設定に影響する要因の整理と、その判断によるデータモデルへの影響の考察などを一般化し、ルールや基準とする必要があると考えている。

次節に、我々の収集した症例をいくつか挙げる。それらをパターン毎に整理し、その対処、判断の基準などを与えている。

#### 5. 症例

例1：キーが複数の意味を包含する

例えば「製番」はあるものを識別するキーである。しかし、それが指し示すエンティティに対する名称はない。また「製番」は製品コード、工程No、設備No、従業員Noなどが入った複合キーのかたまりであり、分解することも可能である。

⇒実務担当者には「製番」が管理単位であるという認識があるので、その製番体系を見直すことがシステム化の目的でなければ、実務担当者の認識を優先する。「製番によって識別されるエンティティ」という抽象的なエンティティを認識する。

例2：同じ識別キーと属性をもつ

「受注」と「受注履歴」は識別キーも属性も同じである。「受注」は代金回収が終わっていないもので、「受注履歴」はいままでの累積である、というような場合。

⇒業務のライフサイクルを描いて検討する。ライフサイクルを考えると、業務として別に扱われていることがわかるので、これは別のエンティティとして認識するべきである。また、同様に同じ識別キーと属性をもつ「編成予算」と「確定予算」では、同じ「予算」というエンティ

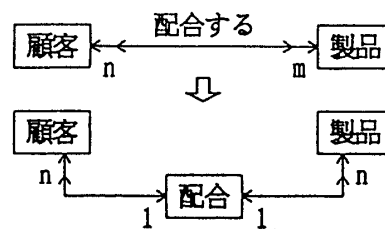


図2 部分的なデータモデル

ティが確定前か後かという業務のタイミングで別物と認識されているだけである、などという場合もある。この場合も、ライフサイクルに着目すると判断できる。

例3：関連のエンティティが先に認識される

「顧客」毎に、「製品」の配合率が異なる為に「配合」というエンティティが先に認識されている。しかし、「配合」には属性が配合率しか考えられず、エンティティかどうかの判断を迷う。

⇒部分的なデータモデルを起こして検討する(図2)。「製品」「顧客」は「配合する」というn対mの関連にあり、n対m関連を1対n関連に分解した場合のエンティティが「配合」とであると判断できる。

#### 6. まとめ

現在、エンティティ認識の問題について、症例を整理したノウハウを収集している。最終的にはこれらの事例からルールや基準を作っていく必要があるが、まず、一つ一つの事例とその対処を収集することが大切であろう。

データ分析では、その企業文化などに依存する部分が大きく、ガイドラインがあっても理屈どおりにいかない部分が多い。実践場面において、企業の実務担当者の考えが不可欠であるが、それを整理するときに、このようなノウハウ集が役立つものと考えられる。

参考文献：

「ソフトウェア構造化技法」1985 J.Martin C.MaClure 國友 他訳 近代科学社

「システム要求分析技法 FUJITSU C-NAP II 解説書」1990 富士通株式会社