

社員番号などのコード非依存のデータベース設計理論

Database Design Theory Unmentioned on Code System

堀田正隆
Masataka Hotta

株式会社プロ・アクシス
ProAxis Ltd.

要旨

現在の情報システムは社員番号、製品番号など、そのものの実体を表す名称とは別に実体を表す符号としてのコードが不可欠なものとして存在し、その存在を前提として情報システムは作られている。データベースもこれに準じて設計されており、情報システムから情報を得るには常にコードが媒介している。しかし、現実の社会（以降、実社会と言う）は実体は名称で識別されている。例えば、人という実体を特定するには「どこそこの何々」と言う。この「どこそこの何々」をもって実社会と情報システムのデータベースを『誤り無く』繋ぐことは不可能であるか？データベースの設計と更新の方式を変えることにより、実社会において、例え「どこそこ」が変わり「何々」も変わっても可能となる。これによって、例えば社員番号を記憶しなければならないとか、製品番号を記憶しなければならないという煩わしさから解放され、企業情報システムをより身近なものとして活用できるようになる。

1. 情報システムにおいてコードが担ってきた役割

これまでの情報システムにおいて、なぜコードが必要であったかを振り返ると、次の6つを挙げることができる。

- (1) 数値にすることでデータ入力の効率性を高める（効率）
- (2) 記憶容量の節約（経済性）
- (3) 名称が違って同じモノは同じに扱う（簡素）
- (4) コードの各桁に意味をもたせることでコードによる分類を可能にする（分類）
- (5) 名称では相当長くなるのでコードで呼び合う（簡便）
- (6) 名称などが変わっても同一実体であることを保証できる（継承）

以上の6つのコードの存在理由のうち上4つの効率、経済性、簡素、分類は否定される。残り2つの存在理由の継承と簡便は引き続き肯定される。この二つのうち実社会においてコードが人目に触れるのは簡便の理由だけで、継承の理由でコードが人目に触れることはない。

先ずコードの存在理由について上4つが無くなった根拠を論理的に述べる。(1)効率と(2)経済性の否定については、2.現在の情報システムの活用環境の章で述べる。(3)簡素と(4)分類の否定については3.コード化の限界で述べる。

次に(5)簡便については具体的な使われ方についてのみ述べる。最後に(6)継承については本論の核心であるので、以降はこれを中心に述べる。

2. 現在の情報システムの活用環境

現在の情報システムの活用環境について、とりわけ大きく変化したことを3つ情報リテラシー、ハードウェア、ソフトウェアの3つに分類して挙げる。先ず、情報リテラシーについては誰もが英字キーボードを扱い可能になったことが過去と比べての大きな変化である。次にハードウェアについては、記憶容量の巨大化、低価格化が過去と比べての大変化と言える。そして、ソフトウェアについては、3つの特筆すべき進歩がある。リレーショナルデータベースの開発、漢字表記の開発、マイクロソフトウインドーズに代表され

るグラフィック・ユーザー・インターフェース（GUI）の開発と普及が挙げられる。こうした発達によってコードの必要性は大幅に縮小した。

先ず、数値を入力することで入力効率性を高める（効率）ことについて言えば、誰もが英字キーボードを扱い可能になったことで数字を入力することも文字を入力することも速度として大差なくなった。むしろ、数字入力の方が視覚的に誤りが気づきにくい分エラー・チェックに時間を要し効率が悪いとも言える。80539でなければならぬものが83509になっていたとしても注意して見ないと気づきにくい。しかし、リンゴでなければならぬのがバナナになっていけば一目で気づく。また、漢字表記の開発とGUIの発達によって数字入力が効率面で絶対的優位でなくなった。現在は、候補名が入力画面上に表示されて、該当するものをワンタッチで選択する方式が可能になっている。なにもキーボードを叩かなくてもデータ入力は可能である。そういう方式が普及している。このように入力データを数値化することで入力効率性を高めることは現在ではほとんど現実性を失っている。

次いで、コードの必要性の第二の理由である記憶容量の節約（経済性）について述べる。記憶装置にバナナと記録するより83509と記録する方がデータ量として小さいことは間違いない。過去において高速に読み書きできる記憶装置は高価なものであり、データ量をいかに少なく済ませるかにはシステム・エンジニアの課題であった。しかし、昨今の記憶容量の巨大化、低価格化によって大半がテキストデータの企業情報システムにおいてはこのことはほとんど問題にならないものとなった。現在ではコードで記録せずに名称で記録したとしても経済性の点で大きな差が生じるとは言えない。

3. コード化の限界

先ず名称が違って同じモノは同じに扱うことについて述べる。同じモノを名称が異なるゆえに別のモノと認識することは企業の効率化を妨げる。同じモノを名称が異なるために別々のモノとして扱えば例えばその分余分な在庫を持たなければならなくなる。こうしたことを防ぐために名称が違って同じモノであれば同じコードを付けて様々なムダを省きたいと考えられてきた。しかし、実社会において異なるという判定は容易だが、同じという判定は実は相当に難しい。AとBが完全に同じモノであると判定することは余程の目利きでなければ難しい。実社会においては「完全に」同じモノであることが立証されなければAの替わりにBを使うことはしない。これは常識と言って良い。従って、名称が違って同じモノは同じコードにするという理想は、判定の困難性ゆえにこれからも遅々として進まないと考える。冷静に考えれば情報システムの担うところから外れている。

次にコードの各桁に意味をもたせることでコードによる分類を可能にするという発想は大変多くの問題を抱えている。その問題を具体的に指摘しておく。

- 1) 実社会には分類が曖昧なものはいくらかでもある。例えば、慶應病院というのは教育機関であるのか医療機関であるのかといったことである。
- 2) 新しい分類上の視点を導入しようとした時にはシステムの作り変えを決断しなければならないほどの困難を伴う。コード体系を変えることとなり、現在使われているコードは使えなくなる。従って、膨大なデータは総書き換えの必要が生じる。
- 3) 変わらない属性などというのはほとんどない。例えば、人間について言えば死ぬまで変わらないのは生年月日と血液型ぐらいで名前はもちろん性別さえも変わる。現在の属性値（例えば性別について言えば男または女）をもとにコードを決めておいたとして属性値が変わったとしてもコードを変えることはできないのでコードによる分類の信頼性は著しく低下する。
- 4) コードによって分類するという発想は分類を徐々に細分化するというツリー構造の発想を前提している。この発想では、上位の桁は大分類、中位の桁は中分類、下位の桁は小分類となる。こうすることによって集計効率を上げている。しかし、実社会においてツリー構造で分類可能というのはごく限られたことであって普通はそのようにはなっていない。哺乳類に性別があるが鳥類にも性別があるというようなことである。

総じて維持の困難性と将来の変更可能性を考えれば、コードによる分類というのは現実的な方法とは言えない。リレーショナルデータベースに任せるべきことである。

本章において、同じモノは同じコードにということとコードによる分類の二点について限界があることを述べた。実社会はコードというものを無条件に受け入れているわけではない。むしろ招かれざる客と受け止めている向きがある。コード化に拘泥することは却って情報システム活用の足を引っ張る懸念がある。その端的な例を示す。新しいモノに対していつでもコード設定できる組織を有している大企業は良いが、そうでない多くの企業は業務に支障を来すことが少なからずある。コードの設定が完了しなければ仕事を先に進められないから困るのである。

4. 名称では相当長くなるのでコードで呼び合う簡便の利点について

ある一つの伝票を厳密に特定するとしたら、発行者（住所と会社名と場合によっては部署名）と発行日時を述べる必要がある。これでは伝える方も骨が折れるし、聞く方も大変である。これらを一言で伝達可能にするには一つ一つの伝票に伝票番号（つまりコード）を振っておくことは有効である。実社会において見聞きする形でコードが存在するのはこの場合のみになる。この場合でもコードは実体を特定する本質ではなくニックネームに相当するものである。本質は発行者と被発行者と発行日(時)であることは実社会において既に確立されている。

5. 継承の意味

人も企業も実体は変わらないのに名称は変わる。モノはモノに名称が付されている限り人や企業のようなことはないが、名称が付されてなければ同じようなことは起る。以下は人や企業について論じる。継承とは、名称が変わっても同じ実体として特定できることを意味する。以下はその事例である。

表1 継承の事例（オリンピックでの個人別メダル獲得数）

年度	種目	階級	メダル種類	選手名	選手番号
1996	女子柔道	48Kg級	銀メダル	田村亮子	53014
2000	女子柔道	48Kg級	金メダル	田村亮子	53014
2004	女子柔道	48Kg級	金メダル	谷 亮子	53014

上の表から、谷亮子のオリンピックでのメダル獲得数を金メダル1個とするのは間違いで、正しくは、金メダル2個、銀メダル1個でなければならない。同じ実体の名称が田村亮子が谷亮子に変わったという事実を伝えているのは53014という選手番号である。このように実体が変わらないのに名称が変わった場合でも、実体として同じであることを伝えるためコードが必要とされる。しかし、JOCも谷亮子本人もこのコードを知っている必要性は全く存しない。本論の本旨であり、これについて以下に述べる。

6. 実社会における実体特定の原則

例えば、企業において同一部署に同姓同名は存しない（紛らわしくならないように人事上そのように配慮されている）。学校において同じ住所に同姓同名の生徒は存しない。しかし、この原則が守られるのはある一時点を捉えてのことであって、異なる時点においてこれらのことは保証されない。つまり、企業を例にとれば表2のようなことが起こらないとは限らない。表2は、同一部署に同姓同名は存しないことの原則に反するように見える。しかし、良く見ると同じ時点で同一部署に同姓同名が存しているわけではないので許される。このことから、次のことが言える。『日時と部署名と氏名を特定すれば必ず唯一、一人

の社員を特定できる』この原則に従えば、人事部も本人も社員番号を知らなくても日時と部署名と氏名で唯一の実体を特定できる。社員番号はコンピュータ・システムのみが知っていて、第一から第二営業部に移動した井上響子が同一実体であることを抑えている。

尚、附言すれば、リレーショナルデータベースの従来のユニークキという概念では十分に覆えなく、この事例の場合では、部署名+社員名+登録日時、の組み合わせでユニーク・キを形成しなければならない。さらに、このことは必要条件であり、同一時点で同一部署に同一社員名が存しないことを確認できて初めて必要十分条件となる。

表2 人事簿の事例

部署名	氏名	登録日時	社員番号	移動日
第一営業部	井上響子	200404021153	53014	20040401
第二営業部	井上響子	200507101812	53014	20050711
第一営業部	井上響子	200510151500	80000	20051013

7. データベースの照会

端末に向かって第一営業部の井上響子を検索すれば、暗黙的に本日の日付が指定されるので社員番号80000の井上響子のデータが提供される。欠勤記録とか出張記録といったトランザクションが人事簿を照会する場合は、欠勤や出張が起こった日付と部署名と社員名を特定して社員簿よりデータを得なければならない。

8. データベースの設計と更新に関するスペック

- 1) データベースの設計においては、移動日(未来日付可)と記録日時の項目が必ず必要である。
- 2) 更新は例外はあるが履歴更新(現在のレコードに上書きするのでなく更新時は必ず新しいレコードを設ける)が原則である。
- 3) 従って、一度作られたレコードは消されることはなく永遠に原状を保持される。
- 4) 新しい者(モノ)には自動的に新しいコードを付与し、引き続きのものにはそのコードを引き継いで、そのコードはデータベースの中でのみで管理し外部に現われることはない。
- 5) 更新時には同一時点で従来の意味のユニークキ(事例では部署名と社員名)の重複が起きてないことを確認しなければならない。
- 6) データベースの照会は従来のユニークキ以外に日時指定が必要である。
- 7) 従って、トランザクション側のデータベースには日時指定できるデータ項目が必須ある(例えば、実施日または記録日時)。
- 8) 欠勤記録とか出張記録はコードではなく名称で記録する。
- 9) 社員簿などのマスターデータが変更されたとしても、それを参照するトランザクションの書き換えは起こらない。

9. まとめ

情報システムを提供する人たちからすると当たり前なことは、実社会の視点でも本当に当たり前のことか。中には暫定的になされたことがいつの間にか恒久化されてしまったこともあるのでないか、いま一度疑ってみる必要がある。本論はコードという問題を取り上げ、情報システムの常識に疑問を投げた。結論は社員コードや企業コードといったものが人目に触れる形で存在する理由は全く存しない。我々の社会からコードを無くすことは可能である。情報システムの常識にはまだまだ疑うべきことが他にも多くある。

今、情報システムに対する企業の期待感は減少しつつある。全てのことについて予断を排して考え直さなければならない時期にある。