

コモディティ化に向かうICTサービスにおけるデータ品質の保証に関する考察

A study of data quality assurance on commoditized ICT services.

羽生貴史[†] 原潔[†]

Takashi Hanyu[†] Kiyoshi Hara[†]

†日本ユニシス株式会社

† Nihon Unisys, Ltd.

要旨

ビジネスの変化に対して柔軟かつ迅速に対応するために、企業内外の複数の情報システムを連携させるニーズがよりいっそう高まってきている。情報システムが多数連携する環境においては、データ品質を高く保つことが今まで以上に重要になるが、実際には毎年のようにデータ品質に関する障害事例が報告されている。本稿ではデータ品質に関する実際の障害事例とデータ品質に関する現行の取組みから、今後の情報システムの環境において、データ品質に関してどのような取組みが必要になるかを考察する。

1. はじめに

今から6年前、ハーバード・ビジネス・レビュー2003年5月号でニコラス・G・カーは“IT Doesn't Matter”を発表し、ITは電話や電力などのインフラ技術と同じようにコモディティ化しており、戦略的価値を持たないと論じた。ダウンサイジングやオープンシステム化、OSS(Open Source Software)やSaaS(Software as a Service)、インターネット・IPネットワークの普及などからハードウェア・ソフトウェア・ネットワークのいずれものコモディティ化しつつあるといえ、これらの技術を使って提供されるICTサービスもコモディティ化に向かっているといえよう。

他方、社会環境や経済環境の変化により、顧客企業のビジネスモデルやビジネスプロセスモデル、情報システムモデルが変化してきている。ビジネス環境の変化に対応するため、意思決定のスピードと質への要求がさらに増しており、それにともない、顧客の情報システムに対する要求も、短期間に変化する傾向が強くなっている。

こうした状況に対し、既存のICT環境を連携させるなどして有効に活用し、いかに迅速に対応するかがこれまで以上に重要になると考えられる。これまで多くの企業では部門毎に個別最適のシステムが導入されていることから、SOA(Service Oriented Architecture)の考え方に則して、再利用しやすい形で既存システムの機能を“サービス”として抽出し、多数の“サービス”が連携して動作することになるであろう。このような環境では、データのコピーが多数の“サービス”間で行われることになり、一旦どこかに間違ったデータが入力された場合にはその修正作業は困難が予想される。

現在、データ品質が起因となる障害事例が毎年のように報告されている。そうした中、今まで以上にシステムが相互に連携する方向へ進展するのであれば、データ品質を保証するためにさらなる工夫を凝らす必要があると考える。

本稿では、2章でIT関連の雑誌記事やIT関連のWebサイトの情報からデータ品質に関する動向を整理する。3章では、データ品質にまつわる障害事例を紹介する。4章では3章の障害事例と2章のデータ品質の取組みを考慮し、今後の情報システムが今まで以上に連携するようになる中で、データ品質に関してどのような取組みが必要になるかを考察する。最後に5章で本発表のまとめを行う。

2. データ品質

2.1. データ品質とは

データ品質とは何か。Wikipediaでは次のように記述されている。「データが高品質であるとは「オペレーション、意思決定、計画などの目的に適っていること」を指す(J.M. Juran)。これとは別に、実世

界の対応実体を正しく表しているデータを高品質とする見方もある。[1]。本稿では前者の定義、即ちシステムで取り扱うデータが特定の目的に適っているかという定義に基づくこととする。情報システムはある目的を達成するために導入されるものであって、その目的を達成しない限り、どんなに忠実に事実を表していようと情報システムに意味はないからである。

2.2. データ品質の歴史的経緯

コンピュータの歴史のごく初期に GIGO (Garbage In, Garbage Out の略。直訳すると「ゴミをいれればゴミが出てくる」。プログラムに組み込まれたロジックに一切間違いがなくとも、与えられたデータ(入力)が誤っていれば、得られる値(出力)は無効なものにしかならないということを示す警句)[2]という言葉が生まれていることから分かるようにデータ品質の保証は古くメインフレームの時代からある大きな課題として知られている。1990年代以降、DWH(Data WareHouse)、SCM(Supply Chain Management) や CRM (Customer Relationship Management) の導入が進み、多数のシステム間でデータのやり取りが必要になる中で注目を浴びてきた。

2.3. データ品質が与える影響

データ品質の低さがどの程度企業に悪影響を及ぼすかについては、以下のような報告がなされている。

- 日本の BI 導入企業の約7割がデータ品質に問題があると認識している。(2008年アビームコンサルティングが発表。連結売上高1000億円以上の東証一部上場企業および有力未上場企業約600社を対象に行ったアンケートより。[3])
- データ品質の問題に起因する損失は平均的な組織で年間820万ドルに上る。22%の企業では2000万ドル、1億ドルを超える企業も4%あった。(2009年米 Gartner 社の調査対象140社のアンケートに基づく TechTarget の記事より。[4])

2.4. データ品質問題に対する業界の取組み

データ品質を保証する上で、これまで一番大きな課題はデータ品質管理が現場任せにされ、経営層の意識が低いというものであった。これについては2.3節で述べたようにデータ品質の低さが経営に与えるインパクトを明確化・定量化したり、SOX法の制定など企業における業務の透明性確保の流れが進んできたことから近年変わりつつある。経営層が理解を示し、データ品質を担保するための専任組織、データ標準化のルール・プロセスの制定(データ辞書・データ標準(データ項目の種類・意味・形式、分散データ間の関係、コード体系などの定義)の作成など)、そしてデータを一元管理するためのIT基盤の構築する企業も出始めている。

一方、ソフトウェアベンダより散在するシステムのデータを有効利用するためにシステム間のデータの整合性を保つためのツールが多数提供されている。ETL ツールに付属するクレンジングツールやMDM (Master Data Management) ツールである。これらのツールはDWHなど情報系システムのDBを構築する際に既存基幹系システムから抽出したデータを正しいものにするために使用されてきている。また、データの発生源である基幹系システムへのデータ入力時に誤入力しないための工夫が多くなされている。UI (User Interface) を工夫したり、上記したツールなどを用いてリアルタイムに入力データの検証処理を行うといったアプローチも多く取られてきている[5]。

3. データ品質に関する障害事例

データの品質が低い(システムで取り扱うデータが特定の目的に適っていない)ことに起因する障害事例は毎年のように報告されている。その一部を整理し、表1にまとめる。

表1 データ品質に関する障害事例

No.	年月	概要	直接的要因
1	2004年4月	Yahoo! ショッピングに出店していた「CatWorks」(カテナが運営)において、アップル製パソコンの価格が2,787円と表示さ	ヤフーの商品登録ミス。販売前にカテナが商品名に対応するJANコードをヤフーに送信したが、ヤフーのシステムで該当するJANコード

		れ、約2万人から1億台以上の注文が入る。	に誤った商品が紐付けられていたため。
2	2005年12月	東証マザーズ市場に新規上場されたジェイコム株に対し、みずほ証券が1円で61万株売り注文を誤発注。直後に3回取消注文を行ったものの、受け付けられなかったため、反対売買を執行したが、9万6236株の買い注文は相殺しきれず、売買が成立してしまった。	誤発注についてはみずほ証券の担当者への入力ミス(「61万円1株売り」とすべき注文を「1円61万株売り」と誤ってシステムへ入力)。3回の取消注文が実施できなかった原因は東証システムの不具合。
3	2007年5月	個人に紐づけることができない不明の年金記録が5000万件以上あることが判明。	入力ミスや改竄を防ぐ仕組みの欠如(2月30日生まれのような異常値を弾く仕組みがなかった、当初漢字だった厚生年金の氏名記録をカナに変換する際にツールで強引に一括変換し、元の漢字データ・変換ツールを削除したなど)。誤入力データを修正する仕組みの欠如(入力データを定量的に把握して改善活動した形跡がない、システム刷新のデータ移行時に見つかった異常記録の修正を20年以上放置など)。
4	2009年10月	イトーヨーカ堂のインターネット通信販売サイトで約4時間にわたり、複数品をセットで販売する商品の価格を、誤って単品価格で表示、販売していた。	イトーヨーカ堂担当者の入力ミス。複数品をセットで販売する商品の価格を、誤って単品価格で表示、販売していた。

4. データ品質を保証するために必要な取組み

3章の障害事例と2.4節の業界の取組みを通して、データ品質に関してどのような取り組みが必要になるかを考察した。今後の情報システムが今まで以上に連携するようになることを考慮した結果、3つの取り組みが必要であると考えられる。

4.1. システム設計時の考慮

当たり前のことであるためなのか、データ品質の問題を語る場合に触れられることがほとんどないが、システム自体が破綻していないように作られる必要があると考える。ここでいう「システム」とは、情報システムのみならず、業務の仕組みという意味も指す。例えば、年金記録の事例では「年金給付の時には本人が来るから、その時に記録と突合して確認し、齟齬があれば直せば良い」という裁定時主義と呼ばれる考え方にに基づき、業務の仕組みが作られ、またそれに合わせた形で対応する情報システムが構築されてきた。そのような「システム」上では、データ品質を保証するためのスキームの計画すら作成されないこととなる。

また、情報システム構築上の根本的な考え方があっていなくても、構築する上で「道具」の使い方を誤るとデータ品質を保つことが難しいことも年金記録の事例は示している。当初漢字だった厚生年金の氏名記録をカナに変換する際にツールで強引に一括変換した件である。言うまでもなく、氏名のフリガナは様ではなく、機械的に正しいものを割り振ることはできない。この状態で元の漢字データと変換ツールの削除をされてしまったら、後から元のデータを修復し、別手法を試すことすら不可能である。

4.2. 妥当性の考慮

データの品質の1つの指標として妥当性がある。非現実的な値が設定されていないかというもので、例えば年齢が200歳と記録されていれば、妥当性はないと判断される。この妥当性の観点でのデータ品質をより向上させる必要があると考える。現在も精度向上を図るための、データ入力時に検証処理をしているが、その処理の高度化が必要になると考える。イトーヨーカ堂の事例の場合には、「仕入価格を参照して利益率を算出し、それが一定の範囲を外れた場合に警告を出す」という比較的単純なルール

による検証処理を追加することが対策になるが、これは現行の技術で対応可能であろう。しかしながら、より複雑なルールで表現される高度な判断が必要なものも情報システムによっては存在するのではなからうか。そのような情報システムにおいて、データ入力時に利用者に不快感を抱かせない形でチェックするためには単なるUIの改善のみでは実現できないと考える。実際、みずほ証券の事例においても、入力チェックの仕組みはあったものの、警告があまりにも頻繁に出るため、利用者が警告を無視していたと言われている。

4.3. (外部・他社との) データ連携時の考慮

Yahoo!ショッピングにおける事例では、カテナは正しいデータを送信したにも関わらず、間違っただデータが表示され、その対応に追われることとなった。データ管理の基本として「One Fact In One Place (1つの事実は1箇所で管理する)」という言葉があるため、この事例のようにシステム連携時にJANコードのように必要最低限のデータのみやり取りすることは一見正当のようにも思える。しかし、これは連携するどこか1箇所でデータを間違えた場合には、連携先のすべてに予想しない挙動が起こる可能性が生じることを示している。連携先が社内に閉じているのであればともかく、自社の管理が及ばない外部との連携をしている場合には、連携先の情報システム内でやり取りしたデータがどのように処理されるのかも確認したうえで、そのフォーマットを決める必要がある。

5. まとめ

入力データを標準化したデータに合わせるようにする、そしてその入力したデータが汚されていないように社内の組織・プロセスを整備するというアプローチが広く浸透しつつあることから、データ品質の問題は徐々に減っていくであろう。しかしながら、データ品質に関する障害を完全になくすためには、まだまだ対応が不十分であるといえよう。今後、本稿で取り上げた取り組みの実証的な評価を通じて、多数の情報システムが連携する中でデータ品質を保证するための仕組みに関する知見を深めていきたい。

参考文献

- [1] “データ品質 - フリー百科事典『ウィキペディア (Wikipedia)』”, <http://ja.wikipedia.org/wiki/%E3%83%87%E3%83%BC%E3%82%BF%E5%93%81%E8%B3%AA>.
- [2] “GIGO (garbage in, garbage out) - @IT 情報マネジメント用語事典”, <http://www.atmarkit.co.jp/aig/04biz/gigo.html>.
- [3] “戦略的ビジネス・インテリジェンス - BI 活用実態調査,” アビームコンサルティング株式会社, 2008, <http://jp.abeam.com/result/RR069.html>.
- [4] Jeff Kelly, “データ品質問題による損失発生を防ぐには - TechTarget ジャパン”, 2009, <http://techtarget.itmedia.co.jp/tt/news/0909/18/news02.html>.
- [5] “特集1：崩れるデータ品質”, 日経コンピュータ, 2007年10月1日号, pp.44-61.
- [6] 水谷哲也, “「eMac」を2787円で販売！ヤフーショッピングの誤表示事件”, 2004, <http://allabout.co.jp/career/corporateit/closeup/CU20040424A/>.
- [7] 東証システムダウン問題取材班, “みずほ証券の誤発注, 問われるフェイルセーフ - ITレポート (動向/解説): ITPro”, 2005, <http://itpro.nikkeibp.co.jp/article/COLUMN/20051227/226805/>.
- [8] “特集1：年金記録問題は防げた”, 日経コンピュータ, 2009年1月15日号, pp.30-45.
- [9] “年金記録問題検証委員会報告書”, 総務省, 2007, http://www.soumu.go.jp/menu_news/s-news/2007/071031_3.html.
- [10] “1個と1箱間違えた? ヨーカ堂通販サイト,” YOMIURI ONLINE (読売新聞), 2009, <http://www.yomiuri.co.jp/national/news/20091021-OYT1T00174.htm> (リンク切れ).