Association for Information Systems
AIS Electronic Library (AISeL)

SAIS 2007 Proceedings

Southern (SAIS)

3-1-2007

Financial Data Quality – A Business Information Systems Perspective

Harry Katzan
katzanh@savstate.edu

FINANCIAL DATA QUALITY
A BUSINESS INFORMATION SYSTEMS PERSPECTIVE

Harry Katzan, Jr.
Savannah State University
Katzanh@savstate.edu

Abstract

The subject of data quality has become increasingly important in recent years as capital markets evolve from supply-based businesses to demand-demand driven enterprises. More organizations understand the impact of poor-quality on business performance and recognize that there is more to the subject than "data cleansing." Service systems based on service-oriented architecture encourage a closer look at the data quality tools, and the impact of quality financial reporting on market value. This paper gives a business perspective on financial data quality from the viewpoint of enterprise systems architecture.

Keywords: Data quality, information framework data exchange

Introduction

The subject of data quality is constantly in the news with cases ranging from lost billing revenues to Internet access. The message is always the same. If information systems were designed properly, used effectively, and the sharing of data were put into place, then there would not be a data quality problem. Moreover, almost everyone these days is familiar with the total data quality mandate and the old adage that *quality is free*. Well, the most knowledgeable and talented people available have applied their talents to the computer field, so that information systems *are* in fact designed properly and modern computer users are more than qualified. With the Internet and all of that, information *is* shared – perhaps more than we would like it to be. We still manage to have data quality problems. Clearly, total data quality in not the answer and data quality is definitely not free. De Guzman (De Guzman , 2007) reports that up to 75% of information workers admitted they have made wrong business decisions because of inaccurate, incorrect, or incomplete corporate information, where "information workers" are employees who use data from various applications, such as spreadsheets, business intelligence reports, and executive dashboards. Miller (Miller, 2002) stresses the futility of operating a business only on supply without regard for demand, and emphasizes that practitioners, academics, and regulators tap into the same economic forces that have aided demand-driven enterprises to succeed in an information-rich environment.

Modern Business Considerations

Financial data quality is as much as an attitude as it is a set of specific practices. If data is incomplete from either an information systems perspective or the market perspective, the following concerns become apparent (Miller, 2002):

- Incomplete information fosters uncertainty
- Uncertainty creates risk
- Risk motivates investors to demand a higher rate of return
- That demand results in a higher cost of capital and lower security prices

To Miller, the solution lies in making better choices when making financial statement; to information systems practitioners, the solution is information completeness.

The Narrow View of Data Quality

It is customary to think of financial data quality as having two components: accuracy and completeness. *Accuracy* refers to whether the data elements are correct or not, and *completeness* involves whether or not it's all there.

Implicit in this view is the concept of the *financial data pipeline*, as suggested by Figure 1, and to some extent, the local definition of data quality is determined by where you are in the chain. However, in spite of the narrowness of this viewpoint, several noteworthy conventions for data handling can be identified:

- There should be a consistent definition of data elements – especially between disparate application systems
- Data should be entered (i.e., placed on computer media) only once.
- Access to data should be regulated.
- New data should be subject to editing, consistency checking, and auditing – all on an automatic basis.

Moreover, the same procedures should be applied to data elements when they are transferred between platforms, as when they are entered initially. Often, a reconciled form of data is held in a staging database between the operational and derived databases. Data quality processing should be performed during reconciliation. On the surface, a narrow view is sufficient, but only if compatible data standards and interoperability are the sole issues. There are other important considerations. Once data is distilled by an application system, then other quality dimensions, such as believability and relevancy and objectivity, come into play. Let's look at the business view.

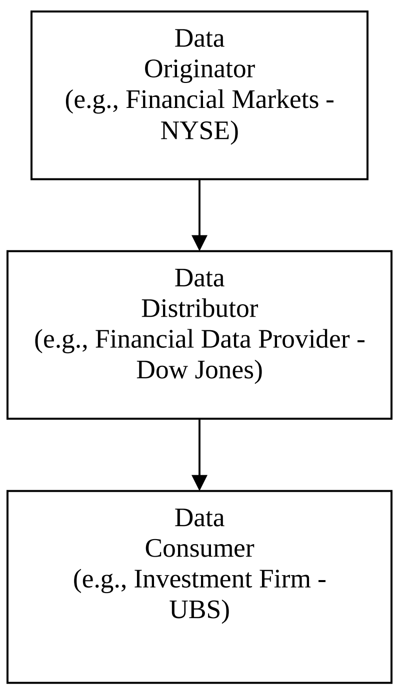

Figure 1. Financial Data Pipeline

A Business View of Data Quality

A business View of data quality would necessarily ask the following obvious question: What do you want to do with the data? Here is a simple list of business goals:

- Better management decisions
- Improved client service
- Increased operations productivity

Admittedly, these are lofty goals, but it would seem that a sound data management program could go a long way towards resolving some of the issues. To make a modest start at the goals, it is necessary to take another look at the attributes usually associated with good data quality. Table 1 gives a more comprehensive list of financial data quality attributes – including accuracy and completeness, given previously. The attributes to some extent parallel the financial data pipeline. The business case for data quality is inherent in two of the attributes, known as business rules and business purpose. The *business rules* attribute refers to derived data and specifically, whether or not enterprise-wide procedures are used in their computation. *Business purpose*, as in "What is the business use of that data?," reflects that a stated business objective is satisfied by the storage and management of a particular data element or set of data elements. It is commonplace for investment firms to calculate and store on a periodic basis a significant amount of data computed using obsolete business rules for historical purposes.

The business perspective is a good segue into the fact that there is a process component to data quality, and it is a subject worth considering.

Table 1. Commonly Accepted Attributes of Good Data Quality

Attribute	*Category*	*Intension*
Accuracy	F	Correctness of information stored as data
Completeness	F	Values for all required attributes are available
Domain Integrity	F	Values of data elements fall within a specified range
Accessibility	O	Ease with which corporate data can be accessed by business units
Consistency	O	Degree to which the meaning of the data element(s) is the same among applications
Integrity	O	Data should not be lost or destroyed
Interpretability	O	Convenience of extracting meaning from the data
Timeliness	O	Degree to which data is up-to-date
Business Rules	B	Data should adhere to specified business rules
Business Need	B	Data should serve a well-defined business purpose

Legend: F – Fundamental, O – Operational, B – Business

The Process Perspective

We are constantly reminded that quality is quality is quality – cars, data, telephone service, and so forth. This is the *conformance to standards* view of data quality, prevalent in modern business that views a quality statement as a hierarchical set of specifications. Take an automobile design as an example. At the general level, a quality specification may include good acceleration, responsive handling, and good fuel economy. At a mid-level, a corresponding set of specifications could include acceleration for 0 to 60 in 6 seconds, stopping distance from 60 to 0 in 190 feet, and a 3.0 liter motor. At a low level, the specification would identify particular part weights and dimensions. But, what about the design? Someone must be thinking about the way a car looks, otherwise we would have a bunch of Edsels running around the streets. There is, of course, an overall design, and in the world of information systems, we would call it a solution. In an information system, the highest level of design is the workflow representing timeliness. The intermediate level would necessarily be the function of the system representing availability. The lowest level would perhaps be the data itself representing accuracy. Evernden (Evernden, 1996) outlines an information framework comprised of three views for managing information: the organization view, the business view, and the technical view. The *organization view* incorporates strategy, organization, and skills. The *business view* concerns data, function, workflow, and solution. The *technical view* encompasses interface, network, and platform. The business view is particularly significant to the areas of financial data quality and data architecture and is outlined here. Within the business view, the *data* refers to the application and historical databases, the legacy files, the staging database, and the transfer files. The *function* denotes the applications, including programs that transfer data between applications. The *workflow* refers to the behavioral aspects of the combination of data and function, including events that trigger activities and the business rules that govern the processes. The *solution* is the data architecture that holds information about the combination of generic components from the data, function, and workflow views. Thus, each time a financial organization develops an information system application, the requirements are based on a combination of data, function, and workflow components that collectively represent the solution. Therein lies a framework for financial data quality based on data architecture.

The Information Framework

A *framework* is a taxonomy for looking at an application domain. It effectively works by identifying the pieces that constitute an application and prescribing how they should be put together. Thus, it is possible to envision frameworks for a variety of industries, such as financial services, manufacturing, engineering, and so forth. Figure 2 gives a reference data architecture for the financial services area. The design emphasizes segmentation and sharing. Operational and reference data are separated, and operational, as well as reference data can shared among applications. For example, operational and historical risk data can be accessed from a portfolio application. The adoption of a framework pertinent to the financial services sector that is based on a reference data architecture provides several salient benefits otherwise unavailable to a financial enterprise. The first benefit is that of *ownership*. Through ownership, a provider-consumer relationship is established between business units resulting in continuous improvement, as suggested by Figure 3. Clearly, this architecture is analogous to the financial data pipeline depicted previously in Figure 1. With this approach, business users inform business owners of the intended use of data and any attendant data quality issues. Inherent in ownership are the “system of record” responsibilities that encompass many of the basic data quality attributes, such as accuracy, completeness, consistency, and timeliness. A second benefit is that of *accessibility*. Through the use of a data hub (or a small number of data hubs), business users can access “system of record” data in a timely and efficient manner. Gone is the use of multiple connections between applications and the high number of transfer files that often result in operational nightmares and a source of inefficiency. A third benefit is the use of a *metadata repository*. Metadata denotes “data about data,” and repository refers to a special database established for the specific purpose of maintaining a high-level architecture. Clearly, a metadata repository takes care of consistency and interpretability and helps with the transfer of data between business data owners and business data users.

The Data Exchange

A *data exchange* is a middleware system designed to transfer data between a source location and a target location. In general, the source and target can be database systems or legacy files. It is normally the case that data exchange programs are hard coded and operate efficiently. An added benefit is that they frequently incorporate data cleansing facilities and data transformation features, which can be an added incentive to employing this approach. In a very general sense, data exchange facilities come in two varieties: those designed to transfer large amounts of data – such as might be expected when transferring data from a data warehouse to a data mart – and those designed to handle small amounts of data on a demand basis. In the former case, the software is known as a *data hub* that works with a metadata repository to achieve high levels of scalability and performance. In the latter case, the software is known as a *data switch* that is designed to transfer real-time requests between applications, such as one might expect, for example, when incorporating risk data from a risk application into a portfolio construction application.

Believability, Relevancy, and Objectivity

In an enterprise environment, people want to know that their data is believable, that it is relevant to the needs and goals of the organization, and that it is collected, stored, and managed in an objective manner. Clearly, this will be the case if the data administrator adheres to a sound data quality program. A sound data quality program binds together many of the ideas presented previously. To start with, it is important to identify the types of data quality metrics that contribute to superior data quality. We have already endorsed the narrow view of data quality that necessarily includes attributes such as accuracy and completeness. But, what if we cannot ascertain that the data is absolutely correct? Well, we can insure that the data is believable and relevant, and this can be done by looking at five forms of integrity:

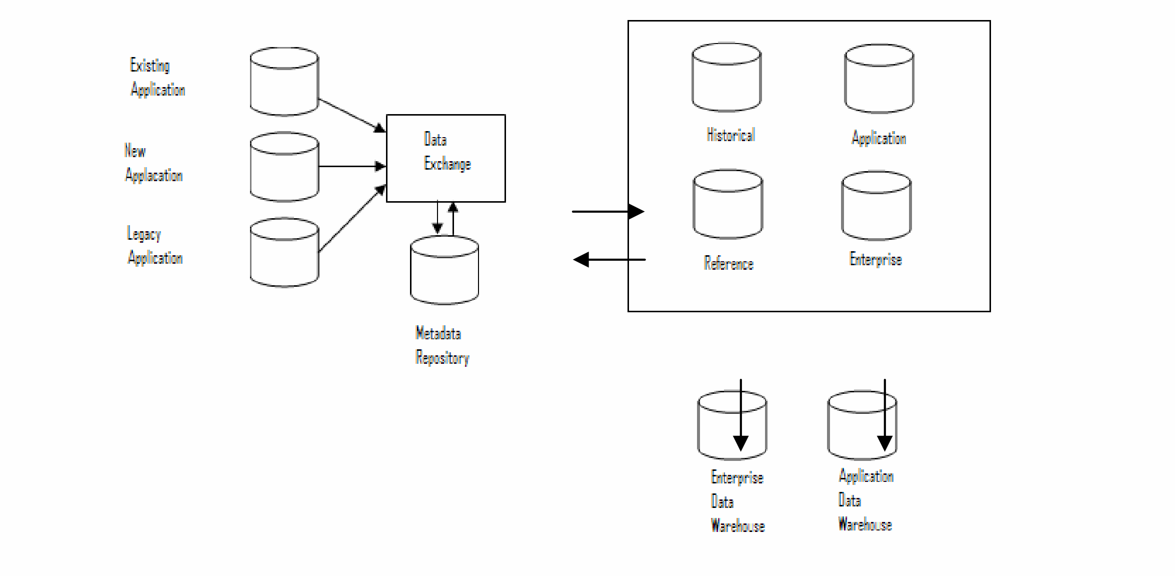

Figure 2. Reference Architecture for Financial Data Quality

- Domain integrity – all values in a column fall into the same predefined domain
- Entity integrity – a primary key is defined for each table in a database
- Referential integrity – each distinct foreign key in a table has a reference in that database
- Column integrity – a constraint applied to the values in a column
- User-defined integrity – specific integrity constraints defined by business rules

Next, we consider four basic actions:

- **Define** – For each type of data integrity, *define* precisely what that aspect of data quality means.
- **Measure –** For each type of data integrity, *measure* the extent to which the data at hand satisfies or does not satisfy the stated data quality definition.
- **Analyze –** For each type of data integrity, *analyze* the measured results to determine the procedures needed to raise the level of data quality to meet enterprise objectives.
- **Improve** – For each type of data integrity, *improve* the quality of the data based on the measurement and analysis.

Finally, we bring business rules into play by applying the various actions to the forms of integrity. For example, the definition of a column integrity constraint for the *gender* column would be that the values must be M or F. Similarly, the definition of a domain integrity constraint for the *customer* column could be that its data type be integer. Huang (Huang, 1999) gives a methodology for data quality that incorporates many of the subjects covered here. One of the basic tenets of good data quality is that a system of record should exist, which is tantamount to saying that business units should own the data and be responsible for its quality.

The Data Quality Tools Market and the Costs

Primarily, the data quality tools market offers stand-alone software products for addressing various aspects of data quality (Friedman, 2006):

- Parsing and standardization to meet standards for business rules and knowledge bases
- Generalized cleansing to satisfy domain restrictions
- Matching to achieve consistency among related data entries
- Profiling to analyze potential data quality issues
- Monitoring to enforce data quality standards for the enterprise
- Enrichment to enhance data from internal or external sources

In an ESA environment, the support of integrated facilities to support integrated ERP, CRM, and BI become increasingly important. The cost of poor data quality can be enormous. Practically everyone is familiar with failure costs versus prevention costs and their internal and external components. But, what about the accuracy and consistency of derived data? It is often been stated that it is better to have consistency than accuracy, but in the modern view of things, we need both. The result is that many project managers prefer to deal with source data and shy away from data produced by other applications. With good data architecture, much of the redundant processing in the financial community can be avoided.

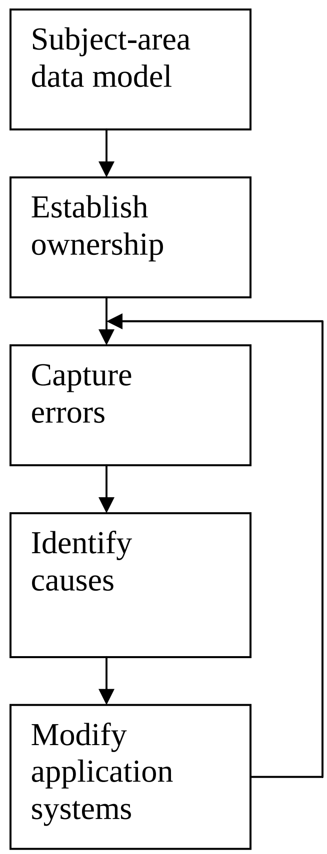

Figure 3. Data Ownership and Continuous Improvement Are Key Elements in Financial Data Quality

References

Devlin, B. (1997) *Data Warehouse from Architecture to Implementation*, Read MS: Addison-Wesley.

Evernden, R. (1996) "The Information Frame Work," *IBM Systems Journal* (Vol. 35, No. 1), 1996, pp. 37-68.

Huang, K. et al., (1999) *Quality Information and Knowledge*, Upper Saddle River, NJ: Prentice-Hall.

De Guzman, M., (2007) "Dirty data puts a smear on efficiency," IT World Canada, February.

Miller, P. (2002) "Quality Financial Reporting," *Journal of Accountancy Online*, April.

Friedman, T and A. Bitterer (2006) "Magic Quadrant for Data Quality Tools 2006," Gartner Group, April 21, 2006.