An Exploratory Study of Data Quality Management Practices in the ERP Software Systems Context

Michael Röthlin

michael.roethlin@iwi.unibe.ch

Abstract: Quality data are not only relevant for successful Data Warehousing or Business Intelligence applications; they are also a precondition for efficient and effective use of Enterprise Resource Planning (ERP) systems. ERP professionals in all kinds of businesses are concerned with data quality issues, as a survey, conducted by the Institute of Information Systems at the University of Bern, has shown. This paper demonstrates, by using results of this survey, why data quality problems in modern ERP systems can occur and suggests how ERP researchers and practitioners can handle issues around the quality of data in an ERP software environment.

1 Introduction – Data Quality in the Enterprise Context

Enterprises need a memory, comprising an accurate, timely, and relevant image of the world around them, in order to operate successfully in the marketplace ([Wa02]). Data are created by organizations in almost every activity, recording the history of actions, driving processes and representing a key input for decisions ([Ol03]).

Data quality (DQ) is a multi-dimensional concept ([WSF95]); commonly used dimensions describing the quality of data are for instance accuracy, completeness, timeliness, or accessibility ([WS96]). DQ literature typically defines high-quality data as data that are fit for use by data consumers; a DQ problem can hence be defined as a difficulty in one or more quality dimensions that makes data unfit for use ([SLW97]). Data errors can be caused by either conceptual (design) or operational problems ([WW96]) and have been categorized into intrinsic weaknesses (data values do not correspond to "real-world" facts), representational weaknesses (data are not intelligible or clear), contextual weaknesses (data are not pertinent to the task of the data user), and accessibility weaknesses (data values are not available or obtainable) ([ADA98]).

In the business context, poor DQ impacts the typical enterprise in many ways ([Re96]). At the operational level, non-quality data lower customer satisfaction, lead to increased cost and lowered employee job satisfaction. At a tactical level, DQ flaws compromise decision-making and reengineering projects. Finally, at a strategic level, insufficient DQ makes it more difficult to define and execute business strategies. Obviously, DQ in enterprise information systems (IS) is key for commercial success.

2 Enterprise Resource Planning Systems and Data Quality

Enterprise Resource Planning (ERP) systems are standardized, integrated, business software applications, which organizations (or their implementation partners) tailor to specific business needs. ERP systems handle recurrent business and administrative tasks such as accounting, inventory and human resources management, purchasing, and sales ([Wa02]).

Looking at a typical ERP system such as SAP R/3, the following classes of data are relevant ([BSS98]): Configuration data, entered or modified during system customization, determine the operation of the ERP system; transaction data are created during work with the ERP system and are typically used only once (that is, in a single business transaction); master data finally are used several times, mostly to simplify and automate transactions. These three types of data can all be affected by DQ weaknesses as presented earlier.

Real-world symptoms for DQ problems in ERP systems are, for instance, high volumes of unpaid invoices (due to wrong or incomplete address information) or disrupted deliveries of goods (the ERP system refuses to process deliveries if the inventory level it has recorded is insufficient, even if there is actually enough physical stock). Adopting Olson's classification ([Ol03]), the sources of potential problems in ERP systems data are:

Initial (manual) data entry: Typing errors stemming from human interaction with the ERP system seem unavoidable; a user may introduce an error for instance by typing a correct value in a wrong field or by entering a typographical error.

Data decay: Certain data types are subject to decay. A project cost planning, for instance, must be reworked periodically to follow project progress. Addresses of suppliers, customers, and employees may change, organizations are split or merged – all these changes in the "real world" of an organization must be reflected in the database of its ERP system.

Moving and restructuring: Data interchanges with other IS can introduce problems in the database of an ERP system. First, during data migration in the implementation phase, existing data of doubtful quality may be transferred from legacy systems into an ERP system. Second, ERP systems can exchange data with other IS in real-time, for instance using Electronic Data Interchange (EDI); again DQ dimensions such as consistency and completeness may be adversely affected.

Use of data: Even if the data in the ERP database are accurate and timely, user interfaces such as entry screens or reports may mislead the user. Furthermore, if critical data items such as materials master datasets are polluted, a multiplication of errors will occur, since erred data values are propagated into system- and user-created transactional data.

Several authors have provided explanations for potential DQ problems in ERP systems. On a conceptual level, ERP systems are accused to be too generic in their structure: ERP systems contain data that are not needed by an organization, and not all required data items will be provided by the ERP software ([Br00]). Moreover, after users have entered data in the ERP system, it absorbs the data and users lose control over the data flow ([Ro99]). On a technical level, integrity checking mechanisms of the underlying relational database system are *not* used in many cases (see [RvdRdG98] for an analysis of SAP R/3); instead, application code of the ERP system is responsible for checking the integrity of data values.

3 A Survey of DQ in the ERP Context

As only few, mostly vendor-driven studies on the topic exist (for instance, [Pr01]), the Institute of Information Systems of the University of Bern has conducted an ERP-oriented DQ survey. Research objective of the study was to determine the relevance and the perceived state of DQ, responsibility for DQ, the influence of diverse factors on DQ, and the use of security measures. A self-administered questionnaire was submitted to IS managers of a random sample of 500 large Swiss companies. The survey was conducted from September 5 to November 10, 2003, and received 125 responses. Selected results of the survey are presented below; details can be found in [Rö03].

Relevance of DQ: Almost 80% of respondents fully agree with the first order role of DQ as a prerequisite for optimal process support provided by an ERP system. Two thirds of respondents agree that operational business data should be cleaned directly in the source system. A state of perfect quality of business data is not considered realistic by almost 80% of respondents; two thirds even deny that a perfect DQ would make sense.

Perceived quality level of business data: On a scale ranging from -3 ("very bad") to +3 ("excellent") and with 0 being "neutral", the comparatively best DQ is found in Human Resource data (mean = 2.06), followed by costing (1.69), supplier master (1.57), and customer master data (1.47). Transaction data are reported at an average level of 1.32. No respondent reports a deterioration of DQ over the past five years.

Responsibility for DQ: The person responsible for DQ is primarily the ERP user, followed by process and line managers. Specific corporate functions such as the "data steward" role described by Brackett ([Br00]) are unknown to or rejected by most respondents.

Importance of selected influence factors: As the results of the enterprise survey show, education (aiming at business and IS process knowledge) is the most important factor influencing DQ, followed by organizational, technical, and personal factors (table 1).

Influence factor	*N*	*Mean*	*Std. Dev.*
1. Insufficient user education	124	2.46	0.70
2. Lack of assigned responsibility for data	124	2.37	0.99
3. Poorly managed ERP implementation project	124	2.28	1.00
4. Lack of overall quality orientation	124	1.81	1.21
5. Insufficient adaptation of ERP to organization's needs	123	1.81	1.12
6. User's lack of experience	124	1.76	1.29
7. User's lack of interest in ERP system and processes	122	1.73	1.14

Table 1: Ranking of factors influencing DQ in ERP systems

Actions for ensuring DQ: A large majority of enterprises customize their user interfaces (frequency = 82%), an operation of tailoring having only slight/moderate impact on the effort required for system maintenance and post implementation ([BHL01]). Respondents further report measures such as agreements on DQ standards with external suppliers (54%) or DQ measurements (52%). However, only a minority of enterprises have implemented formal DQ programs (25%).

4 Conclusion and Implications

As the respondents in the survey confirm, managing DQ remains an important aspect of enterprise IS management, also in the context of ERP systems. The mere fact of using a state-of-the-art ERP product does *not* automatically guarantee a high quality of business data; cases of poor DQ in ERP systems have been cited, with resulting process execution problems. However, in line with other studies (for instance, [Pr01]), data quality is widely seen as satisfactory, even in the light of DQ problems and related quality initiatives; Olson explains this phenomenon with a "high tolerance for errors in primary systems" ([Ol03]).

If ERP DQ issues can be generalized as deficiencies in the mapping of facts about real business processes into ERP data ([WW96]), IS *research* has not presented ways for avoiding these problems, so far. Implementation tools provided by software vendors do not fully bridge the gap; scientific methods and instruments addressing DQ issues are needed.

ERP *practitioners* have to understand that, even if most DQ flaws appear in the form of "usual operational errors," low-quality data can adversely affect business success; on the other hand, ERP software tailoring can and should exclude a very large proportion of possible errors. A formal DQ strategy – defining the role of DQ in business terms – and actions to prevent, detect, and correct DQ errors are hence required:

DQ strategy: Despite of the alleged importance of DQ, only a small minority of companies have implemented formal DQ programs. ERP products such as SAP R/3 contain a large number of data that a particular enterprise will never use. Hence, any DQ management initiative starts with a selection of the business critical data elements needing special care. Since DQ orientation cannot be an objective for its own sake, it must be aligned with corporate (IS) strategies, and DQ projects must be subject to cost-benefit analyses.

Error prevention: On a technical side, critical data must be as clean as possible when entering the system; standard ERP user interfaces should hence be customized to best suit entry processes, with respect to DQ attributes such as completeness or conformity to specific business rules. The business impact of low-quality data must be reduced to a minimum; information product approaches ([WLPS98]) can help to design robust data flows from their entry in the ERP database to internal or external information customers. On a human resources management side, DQ has to enter into the personal objectives and performance evaluation schemes of ERP users. Business process and ERP training must include education on why DQ is relevant in a particular context and how it can be achieved.

Error detection: On a technical side, DQ check routines and reports must be designed, and users must be instructed on how to use them. A DQ management system should assist ERP users in signaling DQ errors and triggering corrective actions. A lifecycle oriented data management must specify checks of data against reality, in order to limit data decay.

Error correction: Specialized staff, such as "super users" or first-level support, must be trained to handle DQ problems and possess the necessary rights to correct critical errors.

As a conclusion, modern technology and the promise of data integration do not exonerate practitioners from actively designing DQ into ERP systems; however, as it has been shown, concepts described in the (data) quality literature can be applied to the ERP context.

References

[ADA98] Abate, M. L., Diegert, K. V., and Allen, H. W. A Hierarchical Approach to Improving Data Quality. 1998.

URL: `http://www.dataquality.com/998abate.htm` [as of 2004-04-26].

[BHL01] Brehm, L., Heinzl, A., and Lynne, M.: Tailoring ERP Systems: A Spectrum of Choices and their Implications. In: *34th Hawaii International Conference on System Sciences*. Maui, Hawaii. January 2001. IEEE Computer Society.

[Br00] Brackett, M. H.: *Data Resource Quality: Turning Bad Habits into Good Practices*. Addison-Wesley. Boston et al. 2000.

[BSS98] Bancroft, N. H., Seip, H., and Sprengel, A.: *Implementing SAP R/3: How to introduce a large system into a large organization*. Manning. Greenwich. Second ed. 1998.

[Ol03] Olson, J. E.: *Data Quality – The Accuracy Dimension*. Elsevier. Amsterdam et al. 2003.

[Pr01] PricewaterhouseCoopers. Global Data Management Survey. 2001.

URL: `http://www.pwcglobal.com/` [as of 2003-10-10].

[Rö03] Röthlin, M.: Datenqualitätsmanagement in ERP-Systemen von Schweizer Unternehmen – Ergebnisse der Umfrage 2003. Arbeitsbericht 156. Institut für Wirtschaftsinformatik der Universität Bern. Bern. December 2003.

URL: `http://www.ie.iwi.unibe.ch/publikationen/berichte/`.

[Re96] Redman, T. C.: *Data Quality for the Information Age*. Artech House. Norwood et al. 1996.

[Ro99] Ross, J. W.: Surprising Facts About Implementing ERP. *IEEE IT Pro*. 1(4):65–68. 1999.

[RvdRdG98] Reind van de Riet, W. J. and de Gruijter, P.: Security Moving from Database Systems to ERP Systems. In: *9th International Workshop on Database and Expert Systems Applications (DEXA'98)*. Pages 273–280. Vienna, Austria. August 1998. IEEE Computer Org.

[SLW97] Strong, D. M., Lee, Y. W., and Wang, R. Y.: Data quality in context. *Commun. ACM*. 40(5):103–110. 1997.

[Wa02] Watson, R. T.: *Data Management: Databases and Organizations*. Wiley. New York et al. Third ed. 2002.

[WLPS98] Wang, R. Y., Lee, Y. W., Pipino, L. L., and Strong, D. M.: Manage Your Information as a Product. *Sloan Management Review*. 39(4):95–105. 1998.

[WS96] Wang, R. Y. and Strong, D. M.: Beyond Accuracy: What Data Quality Means to Data Consumers. *J. of Management Information Systems*. 12(4):5–34. 1996.

[WSF95] Wang, R. Y., Storey, V. C., and Firth, C. P.: A Framework for Analysis of Data Quality Research. *IEEE Transactions on Knowledge and Data Engineering*. 7(4):623–640. 1995.

[WW96] Wand, Y. and Wang, R. Y.: Anchoring data quality dimensions in ontological foundations. *Commun. ACM*. 39(11):86–95. 1996.