

# THE IMPORTANCE OF DATA QUALITY FOR SAP IMPLEMENTATION IN MEDIUM-SIZED ORGANIZATIONS

Dr. Hongjiang Xu, Central Michigan University, [xu1h@cmich.edu](mailto:xu1h@cmich.edu)

---

## ABSTRACT

*Data quality issues are critical for any type of information systems. The purpose of this study is to explore the importance of data quality for SAP implementation. This study involved a case study of a medium-sized organization that had implemented SAP R3 as their ERP system. Knowledge gained from this study has the potential in assisting medium-sized organizations to enhance the quality of the data used in their ERP systems.*

**Keywords:** Data Quality (DQ), SAP, ERP, Information Systems (IS)

## INTRODUCTION

Assuring the quality of data is crucial for any type of information systems. In particular, more and more organizations are implementing ERP systems, which require the ability to obtain and maintain high-quality information. The issue of data quality (DQ) is not new, organizations have both benefited and suffered from the large amount of information made available to them [9]. Data management is not just about clean data: it's also about what data exists and how it can be improved [1]. The new concepts and technologies, such as ERP, SCM, CRM, data mining, data warehousing, have made the DQ issues more complex.

SAP, as one of the leading ERP systems, has gained popularity for not only large organizations but medium-sized organizations as well. In order to help raise the awareness of DQ issues and assist medium-sized organizations' handling DQ in the process of ERP implementation, therefore, the purpose of this study is to investigate the importance of the data quality issues for SAP implementation in medium-sized organizations.

## LITERATURE REVIEW

The total data quality management cycle consists of defining, measuring, analyzing and improving data quality through multiple, continuous improvement cycles [6, 7]. The purpose of data production processes is to produce data for information users. High quality of data means fit for use by those users,

and fitness for use involves multiple dimensions: accessibility, relevancy, timeliness, completeness and accuracy [5, 6, 11].

In the data quality field, some critical factors in DQ have been suggested [2]: understand fully what information quality improvement is and why you are doing it, implement information quality improvement effectively, implementing information quality improvement on the right problem, training and communication, incentives for information quality, management commitment to information quality improvement as a management tool, and managing change.

Other researchers have suggested some important areas and steps that may be taken to ensure DQ. Four steps for the initiation and implementation of a successful system's data quality were recommended [3], these steps are as follows:

- establish data quality position,
- formulate a data quality policy,
- determine objectives and
- obtain management and employee commitment.

Furthermore, six important points in managing data quality were proposed by Segev [8]:

- Establish organizational awareness of the importance of data quality and parties responsible for it.
- Define what we mean by data quality.
- Establish information flow and processes map.
- Identify quality problems and their location on that map.
- Identify technologies and practices that can be used to solve these problems.
- Evaluate the cost/benefit tradeoffs associated with improving the quality of particular data or processes.

Research in DQ suggested that organizations should treat information as a product and follow four principles: understand consumers' information needs, manage information as the product of a well-defined production process, manage information as a product with a life cycle, and appoint an information product manager [10]. Researchers also recommended

organizations establish a data quality program which should include the following: set standards, clearly identify data quality with top-level management, make data quality as a top priority, educate data product suppliers, and teach new data quality skills [4].

## METHODOLOGY

Case study methodology was utilized in this study. A medium-sized organization which has implemented SAP R3 systems was chosen to be the research site. Many ERP and SAP related studies are focused on large-sized organizations; the purpose of this study is to explore the DQ issue in medium-sized organizations. Different stakeholders of the systems in the case organization were interviewed. All case study interviews together with the additional documents obtained from the case study were transcribed and entered into NUDIST, a software package for qualitative data analysis. A content analysis of those documents and interview transcripts was then conducted.

Direct quotations from the case study interview transcripts were used. Quotations from case study interviewees represented their own opinions, perceptions, and experiences regarding particular factors or situations. Therefore, these quotes have the potential to assist readers in obtaining insights into the respondents' understanding of the phenomena.

## FINDINGS

### IS and DQ characteristics

In order to protect the privacy of the case study organization and participants, the case site is referred to as Case F, and the participants are referred to by their positions or roles. In the past, Case F had two separate packages for their IS. It was found that after they installed SAP, they had been making improvements, so now they have the integrated IS including their core business system. The current system in Case F has following features:

- Uses SAP R3 as the ERP system
- Uses an *Oracle* database for financial, payroll and HR
- Uses an Intranet system
- Does not have desktop access to the Internet or external email systems for security reasons

In terms of what the system was able to produce for external reporting, data quality was seen as a priority in Case F. In terms of internal management focus,

data quality was seen as less of an issue, because their internal process was not very complex, and they had very stable staffs that have been able to perform quality controls. Data quality was a priority in terms of continuous improvement efforts so as to produce more useful information on an incremental basis over time in the ERP system.

### Understanding of the Systems and DQ

Knowledge of system linkages was seen as very important. As to where that knowledge should come from, the CFO noted:

*Well, in theory, it should be written down, but in practice, it's the combination of your system knowledge and your organization knowledge.*

The CFO was concerned that the organizations could not improve the quality of the information without actually understanding how the system worked and what was required. That might take a certain amount of time to find out, especially for new staff. Someone might be an accountant and know how SAP worked, but it would not be the same in the new organization.

### Input and Internal Controls

Regarding data input with SAP in Case F, they configured the system so that certain information was mandatory with system checks.

The information producer believed that the input control was the most important stage, because the most important quality control should be at the input stage when the information was entered. They set up certain procedures as well as system constraints to ensure only valid data could be entered.

As to whether the human or system internal controls were more important, one of the senior managers believed that it should be a mixture of both human and system controls. However, he also tended to think that the human aspects were harder to control. The human element that was found to be most critical was *trust*. He thought that no matter what procedures and system controls the organization might have in place, if you were unable to ensure the staff were confident to make the right processing decisions, and they had the trust to do that, then all your other system controls would not matter.

*I've been the auditor. I used to audit against all these controls, and after many, many years, I realized that the most important control is trust. Because I can break any payment systems, I can*

*break any control systems that you give me. And on that basis, the control systems, although fantastic, if you don't have good people who you trust ...*

Therefore, he concluded that it was the individual that would have most impact on data quality. Internal controls would not work without trusted individuals.

### **Top Management, Middle Management and Employee's Commitment to DQ**

Although top management commitment was important for data quality, the CFO believed that they would not have too much impact on actual data quality, because they could not have direct influence.

*They can ask for it, but they can't actually get it themselves. And they would not know whether what they were getting was accurate. Most of the top people, top management, are not trained professionals.*

It then came back to the issue previously discussed before: i.e., trust is the most important thing.

*In our organization, I have the responsibility to make whatever decisions I feel are necessary. And if I think that I need support, I go to senior management.*

The IT Manager saw top management commitment to data quality as very important. However, top management was not seen to perform a critical role in ensuring data quality.

It was believed that the person in the middle, who was responsible for linkages, was the key to information quality. The IT manager stated:

*You can get to the senior managers, who've just seen in front of them the information. They assume that what they have is accurate. So, it is important to get that information right, and it's the person in the middle who does that.*

How to maintain IT staff was seen as a major problem. The employee turnover in some industries is very high, and it may impact the operation of the systems and the information quality. This was particularly true for some IT professionals. Because of the significant pay difference, some IT people came to obtain the necessary training from one organization and then went to another higher paid organization or took a contract job. One of the system's general users noted:

*When we installed SAP three and half years ago, I wasn't here. The team who were here all left within 10 weeks ... they saw the money, and they took it...These people took jobs as contractors, and earned lots more money.*

### **User Involvement**

In regard to whether information producers were asked what users wanted or told users what they needed, the CFO's answer was, "It's a bit of a combination." Furthermore, he added that it was probably more for him to advise users of what they should be managing. The reason for that was because the information producers were normally the few trained professionals on the corporate management side who understood the user's real needs. It also seemed that the CFO in Case F was somewhat opposed to user involvement, based on this reasoning:

*It can't be left to the users, because the users are powerless. The users only get what they are given. The users could require additional information, but they can't actually get that additional information without the person, or the people, the team in the middle, someone responsible for saying: 'Yes, we are meeting statutory requirements, and yes at least asking the question, how can we improve what we are reporting.'*

### **Initial Training**

While training in many organizations is under-resourced, Case F was trying to establish a systematic tool to provide new and existing staff members with a better way to learn new things and to find out the necessary steps and knowledge to handle the system and perform their work.

*We currently have a tool which we are working on to stop this. When a new employee comes in, they will have the tool. It's a very simple intranet-based tool, which hopefully will capture the entire business process starting from which form has to be used to how this system has been configured to how to access reports. New people come into the organization, we continue to have that. A lot of people doing processing may have general business knowledge, but they are not specialized professionals. So we have to explain them, that in SAP you must have a debit and credit. You must have an invoice document and then a separate payment document.*

## Risk Management

Case F set risk assessment based on their information needs. Because the risk of having an external link somewhere in the network is greater than the risk that they are willing to accept in having their core business being corrupted, there are no external links to their core business network.

*For security, we do not have any external links to our network. I have to walk outside my room and logon to a stand alone PC to get my emails. We don't have desktop Internet access... Some background of why we have done that. First of all, they have a minimum problem with viruses. When the 'xxxx' virus came up, the top management was very delighted, because the only infections of the virus were run on stand alone PCs, and it never got onto our network.*

## Teamwork

The Case F stakeholders rated teamwork very highly. They also addressed the importance of communications in being able to create good teamwork.

*The biggest challenge is the strategic view to get everyone aware of why we need to collect this information and what the benefits are. And communication of why we collect the information, and then the communications as to the benefits after we've collected that.*

## CONCLUSIONS

It was clear from the findings of this study that data quality issues are very important in ensuring the success implementation of SAP in medium-sized organizations. Although some of the issues faced by medium-sized organizations are similar to those for large organizations, it should be noted that there are some special issues of particular importance for medium-sized organizations. For example, they may not have as much resources for training as large organizations. They also have difficulty in maintaining qualified employees, which makes it harder to ensure high quality data and smooth

systems operation. However, with commitment from all levels of the organizations, and appropriate controls, training and understanding of the systems and DQ issues, it is possible for medium-sized organizations to successfully implement ERP systems.

## REFERENCES

1. Dubois, L. (2005). Ten steps to total data quality. *Business Intelligence Journal*, spring, 10(2), 57-62.
2. English, L. P. (1999). *Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits*, Hoboken, NJ: John Wiley & Sons, Inc.
3. Firth, C. (1996). Data quality in practice: Experience from the front-line. *International Conference of Information Quality*, (ICIQ '96).
4. Huang, H. T., Lee, Y. W. & Wang, R. Y. (1999). *Quality Information and Knowledge*, Upper Saddle River, NJ: Prentice Hall.
5. Lee, Y. W., & Strong, D. M. (2004). Knowing why about data processes and data quality. *Journal of management Information Systems*, 20(3), 13-39.
6. Lee, Y. W., Pipino, L, Strong, D. M., & Wang, R. (2004). Process-embedded data integrity. *Journal of Database Management*, 15(1), 87-103.
7. Madnick, S., & Wang, R. Y. (1992). Introduction to total data quality management research program. *Total Data Quality Management Program*, MIT Sloan School of Management.
8. Segev, A. (1996). On information quality and the WWW impact: A position paper. *International Conference of Information Quality*, (ICIQ '96).
9. Strong, D., Lee, Y., & Wang, R. (1997). Data quality in context. *Communications of the ACM*, 40(5), 103-110.
10. Wang, R. Y., Lee, Y. L., Pipino, L. & Strong, D. M. (1998). Manage your information as a product, *Sloan Management Review*, 39(4), 95-105.
11. Wang, R.Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4), 4-34.