



Proceedings of GLOGIFT 08
June 14-16, 2008
Stevens Institute of Technology
Hoboken, NJ, pp. 672-681

ENSURING DATA QUALITY IN WORKFLOW MANAGEMENT

Sherry X. Sun* and Wayne Huang**

ABSTRACT

Workflow is an important component in most flexible systems for business process management. As unintentional human errors, such as data entry error, and intentional errors that may constitute a fraud can be introduced through human interaction with workflow systems, ensuring data quality has become a prominent issue in workflow management. This has been echoed by the passage of the Sarbanes-Oxley Act that requires public companies to ensure that their processes, especially those involve financial transactions, to be strictly controlled. However, even with existing built-in workflow controls and rule engines, it is difficult to capture data errors and to detect fraud. To fill this void, we need to incorporate data quality review into workflow systems. This paper creates a framework to determine which workflow step needs data quality review and how data review sampling should be conducted. We first investigate the data quality issues critical in workflow management. We then present an analytical model for allocating data quality review and case sampling in workflows. We use a real world example to demonstrate the proposed method. Moreover, we propose to apply the Statistical Process Control methodology, which has been widely adopted in manufacturing, for monitoring the variations of data quality in workflows. In addition, two different ways to implement the proposed method are discussed.

Keywords: data quality review, workflow management, statistical process control

Introduction

The importance of effectively and efficiently managing business processes has long been recognized (Davenport 1992). As the technology for business process automation, Workflow Management Systems provides a platform for coordinating, controlling, and integrating business processes in a distributed work environment (Stohr and Zhao 2001). Given that a workflow is designed to produce specified information output for a particular customer or market, it is critical to ensure the proper execution of a workflow under the direction of quality information and in conformance to the required business policies, thus creating value for both stakeholders and customers.

Given a well-designed workflow, unintentional human errors, such as data entry error, and intentional errors that may constitute a fraud can still be introduced through human interaction with workflow systems. Those errors can cause improper workflow executions and therefore result to huge economical losses to an organization. For instance, in the case described in (Zur

* Department of Information Systems, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong, sherry.sun@cityu.edu.hk, (Tel) (+852) 2788 8697, (Fax) (+852) 2788 8694

** Director of Information Systems, Prudential Financial Inc., 80 Livingston Avenue Roseland, NJ 07068, wayne.huang@prudential.com, (Tel) 973-548-3888, (Fax) 973-548-4800

Muehlen and Rosemann 2005) a simple data entry mistake with payroll date introduced by a staff member can cause massive business interruption. Intentional error such as insurance fraud can also cause significant financial consequence. It was estimated that more than one of three bodily injury claims from car crashes involves fraud; and 11 to 30 cents of every (property/casualty) claim dollar is lost to fraud (McKendrick 2007). Therefore, ensuring data quality is an urgent issue in workflow management. This has been echoed by the passage of the Sarbanes-Oxley Act that requires public companies to ensure that their processes, especially those involve financial transactions, to be strictly controlled.

In this paper, we develop an analytical framework for ensuring data quality in workflow management. Next, we first review relevant work and then present a model for allocating data quality review and case sampling in workflows. We use a real world example to demonstrate the proposed method. Moreover, we propose to apply the Statistical Process Control methodology, which has been widely adopted in manufacturing, for monitoring the variations of data quality in workflows. Two different ways to implement the proposed method are discussed.

Literature Review

Ensuring data quality has been recognized as an important issue in the IS field for more than a decade (Batini and Scannapieco 2006). Data quality is usually considered as a multifaceted concept concerning with dimensions such as accuracy, completeness, and fitness for use (Strong, Lee et al. 1997; Kaplan, Krishnan et al. 1998; Tayi and Ballou 1998; Lee, Pipino et al. 2004). The literature in the area of data quality mainly focuses on how to assess and improve data quality through either data driven approaches or process driven approaches. The goal of data driven approaches is to analyze and assure the quality of data source in the context of databases. Data cleansing (Hernández and Stolfo 1998), maintaining data integrity (Lee, Pipino et al. 2004), and sample-based data quality estimation all belong to this category.

The process driven approaches consider data as the product from a process. Therefore, rather than focusing on data source, the process driven approaches analyze the process that generates data as the output and possibly propose changes to the data production process for the purpose of data quality improvement. For example, an information manufacturing system model was developed to compute the data quality of so called “information products” along three dimensions, i.e., timeliness, accuracy and cost (Ballou, Wang et al. 1998). As an extension of the information manufacturing system model, the information product map (IPMAP) provides more notations to represent the details on how data is processed. Essentially IPMAP still focuses on data and does not offer a process-centric view since the coordination and sequence of different tasks in a process is missing in the IPMAP model and organizational processes are modeled as a “black-box”.

In accounting information systems, data quality is measured in terms of the certainty that a system is free of specific type of errors and control procedures are often used to prevent and detect specific types of data errors (Kaplan, Krishnan et al. 1998). In the conceptualization of accounting information systems by (Krishnan, Peters et al. 2005), different processes transforming data from one format to another are linked through information flow. We consider this approach closely related to the process-driven data quality control in that this model can help determine the effectiveness of a control procedure for data error detection.

As the significance of business process management has been recognized and the shift of information systems from a data-centric paradigm to a process centric paradigm has emerged, managing data quality and reducing risk due to poor data quality have attracted increased attention in the context of business process management (Zur Muehlen and Rosemann 2005;

Bagchi, Bai et al. 2006). According to (Zur Muehlen and Rosemann 2005) data error is one of the major sources for the risk of business processes. (Bagchi, Bai et al. 2006) propose to borrow the concepts from accounting information systems and to add control procedures in business processes for data error detection and prevention.

At the conceptual level, errors with data can be introduced through inappropriate business process design (Sun, Zhao et al. 2006). Those errors can be detected via data flow verification before the process model is enacted (Sun, Zhao et al. 2006). In a production workflow environment, process data errors can be automatically captured through the implementation of a rules engine (Huang and Stohr 2007). Then the remaining challenge is to control data errors induced by human at the operational level. The analytical method presented in this paper enables data quality control at the operational level. Given that basic idea of applying statistical process control for data quality improvement is roughly suggested by (Redman 1997), this paper discusses the detailed guidelines to apply statistical process control to data quality control in the context of workflow management.

An Illustrative Workflow Example

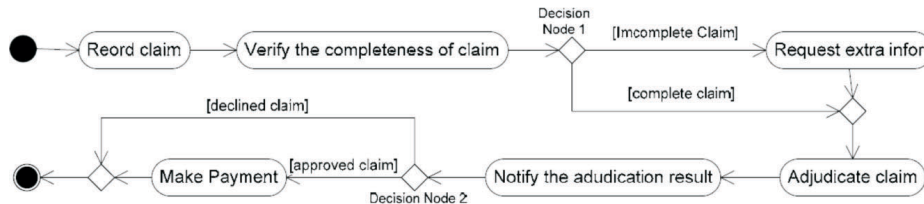


Figure 1: A Workflow Model for Insurance Claim Processing

Throughout this paper, we use a life insurance claim processing workflow as the running example. Figure 1 shows the simplified workflow model in the notation of UML activity diagram. The workflow is triggered when a claim document is received. Then the claim is recorded into the workflow system and verified for its completeness. If some required information is missing, the client is asked to provide the missing information. Then the claim will be adjudicated and the client is informed with the decision. In the case the claim is approved, a payment is made to the client. Next, we use this example to examine the data quality issues in management.

Data Quality Control Issues in Workflow Management

Many data quality concerns, such as data accessibility, data understandability, data fitness for use, and so on, are no longer issues in workflow management since workflow systems can automatically route task along with associated information to appropriate workflow workers. Typical workflow systems can apply data check or process rules to prevent obvious data error or process error (Huang and Stohr 2007), e.g., detecting the claim incurral date is earlier than the policy effective date in insurance claim processing. However, even with various data check and process rules, it is difficult to detect and capture most unintentional and intentional data errors introduced by people through their interactions with a workflow system. Thus, error detection posts a serious challenge for data quality control in workflow management. As a major effective means for detecting and correcting data errors in workflows in addition to incorporating data check rules into workflow engine, data quality review can be applied at certain steps in a workflow and all the critical data involved in the workflow can be sampled and scrutinized for data errors. Given that data quality review requires resources such as labor and time, it is cost-effective to minimize the number of steps where data quality review deems necessary.

In order to decide the steps most appropriate for adding data quality review, we analyze the influence of workflow structures on data quality control. A workflow contains many decision nodes where a proper execution path is determined. For example, the model in Figure 1 contains two decision nodes. Given that the goal of data quality control is to assure proper workflow execution, the quality of data used as input at those decision points is critical. Those data are referred to as decision variables in this paper. Moreover, the activities that generate those decision variables as output are referred to as decision making activities. The result of a decision made by a decision making activity is manifested via a set of decision variables. Once a decision is made, a workflow is executed in accordance to the decision. Thus, the data quality in a workflow can be decomposed into three parts: 1) the correctness of the data that a decision is made upon; 2) the correctness of a decision, i.e., the decision is unbiased and conforms to the fact and the relevant business policies in a business context; and 3) the correctness of the data that the execution of the decision depends on. Next, we use the examples from the insurance claim processing workflow shown in Figure 1 to explain.

Example 1 At Decision Node 1 in Figure 1, if a claim is complete, then the step of *request for extra information* is skipped. Otherwise, a request is sent to the customer asking for the missing information. The decision variable used as input by Decision Node 1 is “claim-completeness-status”, the value of which can be either “complete” or “incomplete”. This decision variable is the output of the step, *verify the completeness of claim*, which is considered a decision making activity. Data quality in this example is determined first by the quality of the input data for *verify the completeness of claim*, which are the foundation for making a correct decision. Moreover, the value of the decision variable “claim-completeness-status” should conform to the real fact. That is, *claim-completeness-status*=“Yes” holds whenever the claim contains complete information and *claim-completeness-status*=“No” holds otherwise. Finally, when the decision leads to an execution of some workflow steps, the input data used by those workflow steps need to be correct. For instance, in the case that a claim is not complete, a request for missing information will be sent to the customer. The data input for the activity *request for extra info*, such as the contact of the customer, needs to be correct for a proper execution of *request for extra info*.

Example 2 At Decision Node 2 in Figure 1, if a claim is declined, the workflow ends. Otherwise, a payment needs to be made. The decision variable, “claim-adjudicationstatus”, can have the value of either “approved” or “disapproved”. The decision making activity in this case is *adjudicate claim*, the activity producing “claim-adjudicationstatus” as output data. Data quality in this example is decided by the correctness of the input data for activity *adjudicate claim*, e.g., date of death and policy effective data, the conformance of the decision to the business policies relevant to claim approval, and the correctness of the data, e.g., beneficiary name and address, that are used as input when a payment is made.

An effective data quality review should help to capture errors and make sure decisions are made correctly and executed accurately. Given that the decision nodes are the points where a decision has not been executed and the relevant decision variables have been produced, decision nodes in a workflow are the most suitable steps in a workflow for applying data quality review. Next we present an analytical model to help design data quality review for a given workflow structure.

A Model for Data Quality Review

Given a workflow with n decision nodes, our data quality review model focuses on determining at which decision node(s) to apply data quality review, how to sample the workflow cases and

monitor data quality in a workflow.

Designing Data Quality Review

A carefully designed integration points of data quality review in a workflow can minimize the cost caused by data errors and the expense associated with data quality review.

Cost of Errors

In a workflow, the cost caused by data errors include the financial cost resulting from the execution of an incorrect decision due to data errors and the resource cost used to execute the incorrect decision, i.e.,

$$C_{error} = C_{financial} + C_{resource}. \quad (1)$$

Assume a workflow contains decision nodes and it takes p_i time units for a resource (e.g., workers) to complete activity i and the cost rate per time unit for the recourse needed for completion of activity i is w_i . At the decision node d , the cost of resources used to execute an incorrect decision can then be calculated as follows.

$$C_{resource} = \left(\sum_{j=1}^{\#_of_paths_endActivity} \sum_{i=id_j} w_i p_i \right) \frac{1}{\#_of_paths}, \quad (2)$$

where id is the first activity to be executed after the decision node d when an incorrect decision is made and $\#_of_paths$ is the number of different paths a workflow can take under the incorrect decision. For instance, in *Example 1* above, if a claim is complete and the decision variable *claimcompleteness-status* is incorrectly set to be “No”, the financial cost of the decision is the missing service penalty if the claim payment is delayed and the resource cost is the average cost of labor time and other resources to finish the entire workflow starting from the activity *request_extra_Info* to when a claim is approved or declined.

Effectiveness of Data Quality Review

The effectiveness of data quality review can be measured through the cost saved from the detection and correction of data errors when data quality review has been applied at a decision node d .

$$C_{saved} = \Delta C_{error} = (1-r) C_{error}, \quad (3)$$

where r is the probability that a data error is not captured after data quality review has been applied at decision node d .

Where to Apply Data Quality Review

Data quality review requires resources such as labor time. Given limited resource that can be allocated for data quality review, the goal of data quality review is to maximize the economical return of error detection while minimizing the cost spent on data quality review. Assume a workflow contains n decision nodes and it takes $(preview)i$ time units for a resource (e.g., workers) to complete data quality review at decision node i and the cost rate per time unit for data quality review is w_{review} . The problem of deciding where to apply data quality review in a workflow can be formulated as follows:

Ensuring Data Quality in Workflow Management

$$\begin{aligned} & \max \sum_{i=0}^n (C_{saved})_i x_i \\ & \text{subject to } \sum_{i=0}^n w_{review}(P_{review})_i x_i \leq limit \\ & x_i = 1 \text{ or } x_i = 0 \end{aligned}$$

When $x_i=1$, data quality review is applied at decision node i ; $limit$ is the expense limit for data quality review of one workflow case.

Dynamic Sampling Method

After deciding where to apply data quality review, rather than applying data quality review for every workflow case, a selective approach can be used to determine an appropriate number of workflow cases for data quality review, to minimize the cost incurred by data quality review. The percentage of the workflow cases receiving data quality review is referred to as sampling rate. According to Sarbanes-Oxley audit practice, process auditors typically set a minimum sampling rate for a set of workflow cases over a period of time, e.g., one month. Furthermore, some workflow cases may have higher financial impact and thus are subject to higher sampling rates. Various sampling criteria can be applied to guarantee the workflow cases with high possibility of data errors and critical financial impact can have more representatives in the samples for data quality review. The sampling criteria can be grouped into two types. The first type of sampling criteria is based on the characteristics of workflow cases that require special consideration for data quality review. For example, insurance claims handled by a new caseworker are likely to have more data and/or decision errors than claims handled by experienced caseworkers. Thus, a certain percentage of insurance claims handled by new caseworkers will be selected for data quality review in the insurance processing workflow. The other type of sampling criteria emphasizes on the workflow cases that are exceptions to business policies. For example, increasing benefit amount after a claim is approved is unusual and requires data quality review. Based on the economical consequence, some sampling criteria may maintain a fix sampling rate while others may have a changeable sampling rate. For example, declining a claim is a critical decision and requires 100% sampling rate. Table 1 provides examples of real world sampling criteria used to select a claim for data quality review in the life insurance claim processing workflow shown in Figure 1.

Table 1 Examples of Sampling Criteria and Sampling Rate

Sampling Criterion	Required Sampling Rate
<i>Special Considerations</i>	
Claim status	100% data quality review for all declined claims
Product type	60% of accidental death claims are subject to data quality review
Specific caseworker	All new caseworker's cases are subject to 50% data quality review
Special client	40% for sensitive account cases
High dollar claim	All claims exceeding \$250,000 are subject to 100% data quality review
<i>Exceptions to Business Policies</i>	
Increase benefit amount after a claim is approved	100% data quality review
Change payee address for a pending	100% data quality review

It is important to design a sampling approach that can ensure the overall sampling rate meets the minimum requirement set by process auditors and various special control standards set by the process owner. On the other hand, we need to minimize the data quality review cost and avoid unnecessary sampling. A dynamic sampling algorithm is developed to solve the challenge

as follows. Assume during the time interval from the beginning to t , $s_i(t)$ is the actual number of workflow cases that match sampling criterion i among the total of $S(t)$ workflow cases being processed within t , $r_i(t)$ is the sampling rate for sampling criterion i at time t , and $p(t)$ is the probability of a claim being sampled a time t .

$$p(t) = \frac{\sum_{i=1}^n r_i(t) * s_i(t)}{S(t)} \tag{4}$$

Assume time T is when the sampling period ends, the objective is to adjust $p(t)$, when $t < T$, to ensure the overall sampling rate at $p(T)$ approaches the proximity of the target sampling rate P . The proposed approach is to adjust the changeable r_i when t is at $\frac{1}{4} T$, $\frac{1}{2} T$, and $\frac{3}{4} T$. For example, at $t = \frac{1}{2} T$ we can find out $s_i(t)$, $S(t)$, and calculate $p(t)$ based on the known $r_i(t)$ with Formula (4) above.

If $p(t) < P$, then increase r_i , if $p(t) > P$, then decrease r_i . For non-changeable r_i , its sampling rate stays the same throughout the entire T period. Assume $s_i(T)$ is the projected number of cases that match sampling criterion i among the estimated total number of cases $S(T)$ processed at time T . The amount of increase or decrease for changeable r_i at time t is $\Delta r(t)$.

$$\Delta r(t) = \frac{P * S(T) - \sum_{i=1}^n r_i(t) * s_i(T)}{\sum_{i=1}^n s_i(T) - s_i(t)} \tag{5}$$

$$r_i(t+1) = r_i(t) + \Delta r(t) \tag{6}$$

Example 3. The following is an example of this approach that uses Table 1 sampling criterion and sampling rates. Assume that the averaged total claim population $S(T)$ in a month (time T period) is 20,000 claims and the target sampling rate P is 10%. Half way through the month ($t = \frac{1}{2} T$), based on the sampling rate $r_i(t)$ listed in Table 2, the actual sample size of selected claims is 1,217 (See Table 2 *Sample Size* column). Assume that the actual claim population $S(t)$ at time t is 9,500 claims; this would make the averaged sampling rate $p(t) = 12.8\%$ based on Formula (5), which is greater than the 10% target rate. Since we cannot change sampling rate for criteria with fixed sampling rate (e.g., those that are having 100% sampling rate), the only criteria for which we can change the sampling rate are: accidental death claims, new caseworker cases, and sensitive account cases. The calculated $r_i(t)$ is -17% based on Formula (6). The new sampling rates $r_i(t+1)$ for the three criteria are listed in Table 2. With the projected population for each criterion $s_i(T)$, the calculated total sample size is 2,000, which meets the target sampling rate 10%.

Table 2 Sampling Rate Change Example

Sampling Criterion Example	Sampling rate % $r_i(t)$	Population for Each Criterion $s_i(t)$	Actual Sample Size	Sampling rate % $r_i(t+1)$	Projected Population $s_i(T)$	Calculated Sample Size
Special Considerations						
Claim status is denied	100%	300	300	100%	500	500
Accidental death claims	60%	500	300	43%	800	429
New caseworker cases	50%	1,000	350	33%	1,500	615
Sensitive account cases	40%	500	160	23%	750	241
High dollar claim	100%	100	100	100%	200	200
Exceptions to Business Policies						
Increase Benefit amount after a claim is approved	100%	3	3	100%	8	8
Change payee address for a pending payment	100%	4	4	100%	7	7
Total			1,217		3,765	2,000

Monitoring and Improving Data Quality

In this section, we discuss how to monitor and control data quality variations in a workflow. We propose to borrow the method of Statistical Process Control to help us focus on data quality problems where significant quality variance occurred. Next, we use the example of life insurance claim processing to explain the application in workflow management.

Table 3 provides examples of claim quality attributes and quality point deduction assignment. After a workflow case is selected for quality review, it is given a quality score based on the number of points deducted from the total score of 100. These quality scores can be averaged at caseworker level and at departmental level, and used as a benchmark for quality improvement. The quality result can also be assessed by product line or by customers.

Table 3 Example of Claim Quality Attributes and Quality Points

Claim Quality Attributes	Quality Points Deduction
<i>Decedent Data</i>	
Claimant Name	2
Claimant SSN	2
Claimant Address	2
First Notice Date	2
Date of Death	2
<i>Beneficiary Data</i>	
Beneficiary Name	2
Beneficiary Address	2
Beneficiary SSN	2
<i>Decision Quality</i>	
Correct claim decision	5
Correct claim benefit amount	5
Payment to the correct person	5
Correct payment to the person	5
Correct interest calculation	5
Correct payment method	5
Correct tax notice	5
Have necessary documentation	5

Note: each claim starts with 100 quality points

Using quality scores assigned to claims, we can apply the Statistic Process Control methodology as the tool for monitoring and improving data quality. The u control chart, which monitors the number of defect per unit sample in each different size samples (Lober 1994), can be applied to workflow quality review process. A defective case in workflow is defined as a case having quality score that is less than the target quality score set by the management. Assume the daily number of cases sampled for quality review on day t is S(t); and the number of defects on day t is c(t), then the average defects per case is u(t) on day t.

$$u(t) = c(t) / S(t) \tag{7}$$

The average defects per case from day 1 to day t is k(t).

$$\bar{u}(t) = \frac{\sum_{i=1}^t c(i)}{\sum_{i=1}^t S(i)} \tag{8}$$

A u chart shows daily u(t) variation. We can further obtain the 3σ lower control limit LCLu (Lober 1994).

$$LCLu(t) = \bar{u}(t) - 3\sqrt{\hat{u}(t)/S(t)} \quad (9)$$

We do not consider the 3δ upper control limit since higher quality score is always better. For a given period of time, e.g., one month, if any $u(t)$ is less than the $LCLu(t)$, investigation should be made to identify the causes of error and corrective actions should be taken. Furthermore, issues that caused these defects can be added to data quality review sampling criteria to ensure the corrective action is effective.

Implementing Data Quality Review in Workflow Management

Data quality review process should be an integral part of workflow design where its outputs have significant impacts to the operational and financial results. Data quality review can be implemented in two different ways as detailed below.

Incorporating data quality review into the existing workflow In this approach, after determining at which steps data quality review needs to be applied, data quality review is added as steps into the existing workflow and the task will be performed by a different caseworker to maintain independent judgment. This design has the advantage of utilizing the existing caseworker resource pool, but it lacks the consistency required by the quality review standard. Additionally, when changes with data quality review are in need, the workflow model needs to be changed accordingly. This approach is more suitable for low volume workflow environment where the resource is limited and the change of quality review process is infrequent.

Operating data quality review as a separate workflow parallel to the existing workflow In this approach, the data quality review is operated as a workflow parallel to the existing workflow, which sends the workflow cases to the data quality review workflow and then wait for the result at the step where data quality review is required. A separate quality review team who report directly to management is usually established, which adds to the workflow operational cost. This design has the advantage of an independent team that can review cases in fair and consistent manner. It is more adaptive to the changes with data quality review and the exiting workflows. This approach is more suitable for organizations with high volume workflow environment where there is enough resource to support the quality review activity and the workflow change is frequent.

Conclusions

Data quality in workflow management is an important issue that requires organization attention to ensure the execution of workflows can bring expected value to stakeholders. Only with quality data can a workflow operates effectively. This is also critical for many flexible systems with workflows as their core components. In this paper we propose a process-centric data quality control that can be implemented in workflow systems. Our contribution includes cost-driven allocation of data quality review, dynamic workflow case sampling, and application of statistic process control to monitor data quality variations in workflows. This paper only presents the preliminary results. In the next step, we will investigate the feasibility of a intelligent real-time data quality feedback system integrated with business process rules engine and then adapt existing data mining techniques for identifying the unknown causes of data error.

References

- Bagchi, S., X. Bai, et al. (2006). Data Quality Management using Business Process Modeling. *IEEE International Conference on Services Computing, Chicago, IL.*
- Ballou, D., R. Wang, et al. (1998). "Modeling Information Manufacturing Systems to Determine

Ensuring Data Quality in Workflow Management

- Information Product Quality." *Management Science* 44(4): 462-484.
- Batini, C. and M. Scannapieco (2006). *Data Quality: Concepts, Methodologies and Techniques*, Springer.
 - Davenport, T. H. (1992). *Process Innovation: Reengineering Work Through Information Technology*, Harvard Business School Press.
 - Hernández, M. A. and S. J. Stolfo (1998). "Real-world Data is Dirty: Data Cleansing and The Merge/Purge Problem." *Data Mining and Knowledge Discovery* 2(1): 9-37.
 - Huang, W. and E. A. Stohr (2007). Design and Implementation of a Business Process Rules Engine. *International Conference on Design Science Research in Information Systems and Technology*. Pasadena, CA.
 - Kaplan, D., R. Krishnan, et al. (1998). "Assessing Data Quality in Accounting Information Systems." *Communications of ACM* 41(2): 72-78.
 - Krishnan, R., J. Peters, et al. (2005). "On Data Reliability Assessment in Accounting Information Systems." *Information Systems Research* 16(3): 307-326.
 - Lee, Y. W., L. Pipino, et al. (2004). "Process-Embedded Data Integrity." *Journal of Database Management* 15(1): 87-103.
 - Lober, S. J. (1994). *Statistical Process Control*. United States Army Depot Systems Command.
 - McKendrick, J. (2007). Sniffing Out Claims Fraud Before the Check is Cut. *Insurance Networking News*: 6-14.
 - Redman, T. C. (1997). *Data Quality For The Information Age*, Artech House Publishers.
 - Stohr, E. A. and J. L. Zhao (2001). "Workflow Automation: Overview and Research Issues." *Information Systems Frontier*.
 - Strong, D. M., Y. W. Lee, et al. (1997). "Data Quality in Context." *Communications of ACM* 40(5): 103-110.
 - Sun, S. X., J. L. Zhao, et al. (2006). "Formulating the Data-Flow Perspective for Business Process Management." *Information Systems Research* 17(4): 374-391.
 - Tayi, G. K. and D. P. Ballou (1998). "Examining Data Quality." *Communications of ACM* 41(2): 54-57.
 - Zur Muehlen, M. and M. Rosemann (2005). Integrating Risks in Business Process Models. *16th Australasian Conference on Information Systems*, Sydney.