Normal Accidents: Data Quality Problems in ERP-Enabled Manufacturing

LAN CAO, Old Dominion University
HONGWEI ZHU, University of Massachusetts Lowell

The efficient operation of Enterprise Resource Planning (ERP) systems largely depends on data quality. ERP can improve data quality and information sharing within an organization. It can also pose challenges to data quality. While it is well known that data quality is important in ERP systems, most existing research has focused on identifying the factors affecting the implementation and the business values of ERP. With normal accident theory as a theoretical lens, we examine data quality problems in ERP using a case study of a large, fast-growing multinational manufacturer headquartered in China. Our findings show that organizations that have successfully implemented ERP can still experience certain data quality problems. We identify major data quality problems in data production, storage and maintenance, and utilization processes. We also analyze the causes of these data quality problems by linking them to certain characteristics of ERP systems within an organizational context. Our analysis shows that problems resulting from the tight coupling effects and the complexity of ERP-enabled manufacturing systems can be inevitable. This study will help researchers and practitioners formulate data management strategies that are effective in the presence of certain "normal" data quality problems.

Categories and Subject Descriptors: H.4.2 [**Information Systems Application**]: Types of System—*Enterprise resource planning*; K.6.4 [**Management of Computing and Information Systems**]: System Management—*Information quality*

General Terms: Management, Human Factors, Performance, Reliability

Additional Key Words and Phrases: Data quality, enterprise resource planning, ERP, normal accident, tight coupling, complexity

ACM Reference Format:
L. Cao and H. Zhu. 2013. Normal accidents: Data quality problems in ERP-enabled manufacturing. ACM J. Data Inf. Qual. 4, 3, Article 11 (May 2013), 26 pages.
DOI: http://dx.doi.org/10.1145/2458517.2458519

1. INTRODUCTION

Many organizations have implemented Enterprise Resource Planning (ERP) systems to support their operations and competitive strategies [Sirkisoon and Shepherd 2002]. An ERP system is often used to replace many separate systems that serve different functional areas such as production planning, purchasing, manufacturing, sales distribution, accounting, and customer service [Belotsky and Johns 2008]. With a centralized data repository and a collection of configurable modules, ERP integrates enterprise transaction processing to balance demand and supply. It also provides useful reports and facilitates data analysis and decision making.

Authors' addresses: L. Cao (corresponding author), Department of Information Technology and Decision Sciences, College of Business and Public Administration, Old Dominion University, Norfolk, VA 23529; email: lcao@odu.edu; H. Zhu, Department of Operations and Information Systems, University of Massachusetts Lowell, MA 01854.

DOI: http://dx.doi.org/10.1145/2458517.2458519

Among many benefits claimed for ERP systems, improved data quality and information sharing within an organization are reported [Xu et al. 2002]. However, due to its complexity, ERP also poses challenges to data quality [Gattiker and Goodhue 2004; Xu et al. 2002]. Although it is well known that data quality is important in ERP systems [Beard and Sumner 2004; Gattiker and Goodhue 2005], ERP-related data quality issues have not been systematically investigated. Most existing research has focused on identifying the factors affecting the implementation process and the business values of ERP (e.g., Somers and Nelson [2004]). In this article, we examine ERP-related data quality problems using a case study of a large, fast-growing multinational manufacturer headquartered in China. Our research questions are: What are the data quality problems in ERP-enabled manufacturing systems? What are the root causes of these data quality problems?

We answer these questions using a case study with the *Normal Accident Theory* (NAT) as a theoretical lens. Developed in research on risks and accidents of large engineering systems, NAT has been used to explain why accidents are inevitable if the systems are tightly coupled and their components have complex, hidden interactions [Perrow 1999]. In other words, it is normal to have accidents in such systems. Our study extends the application of NAT to the analysis of data quality problems in ERP-enabled manufacturing.

The rest of the article is organized as follows. Section 2 reviews the existing research on data quality and ERP-related data quality issues; it also presents the normal accident theory as a theoretical lens to investigate these issues. Section 3 describes the case study method used in this research. Section 4 presents our results of the case study. Section 5 discusses the challenges and potential solutions to ERP-related data quality problems. Section 6 concludes the work and discusses directions for future research.

2. THEORETICAL BACKGROUND

The adoption of ERP systems is often driven by business objectives, many of which are related to the quality of data used for operational and decision-making purposes. In this section we first discuss existing research on data quality and ERP-related data quality problems. Then we discuss the normal accident theory, which is used in this research to examine data quality problems in ERP systems.

2.1. Data Quality

Existing data quality research has identified as many as 16 dimensions to describe various aspects of the general notion of data quality [Wang and Strong 1996]: accuracy, objectivity, believability, reputation, accessibility, access security, relevancy, timeliness, consistent representation, completeness, value-added, interpretability, amount of data, ease of understanding, and concise representation. Certain causal relationships among data quality dimensions have been identified through case studies in organizations [Strong et al. 1997]. For example, when there are multiple data sources for the same data, inconsistency from these sources can cause low believability in the data, which can further lead to the perception of low value-added in using the data, and in the end the data not being used.

A data quality problem can be defined as “any difficulty encountered along one or more quality dimensions that renders data completely or largely unfit for use” [Strong et al. 1997]. To solve data quality problems, an individual’s knowledge on data quality is important [Lee and Strong 2003]. Many practitioners solve data quality problems by crafting preexisting rules to integrate business process and data process [Lee 2003]. They use their context knowledge on data collection, storage, and usage to develop a solution of a given data quality problem. However, it is not clear how to explicate the context of data systematically. Research shows that the effective method of managing data quality is to treat data as a product and eliminate the root causes of data quality problems by improving the data processes [Ballou et al. 1998; Strong et al. 1997; Wang 1998; Wang et al. 1998]. In this research, we try to identify the root causes of data quality problems in ERP by examining organizational context and ERP characteristics.

2.2. Data Quality in ERP Systems

Despite the expectations of using ERP systems to improve data quality in organizations, the impacts of ERP systems on data quality are not well understood. There are only a few studies that attempt to address this issue. A longitudinal case study of a multinational industrial equipment manufacturer shows that the use of an ERP system across multiple sites of the company has increased the overall data accuracy, but the quality in the other dimensions may increase or decrease depending on the user groups [Strong and Volkoff 2005]. For example, certain users may need to supply additional data not related to their tasks, reducing data relevancy to these users, but the data are useful to other users. Common data representation and definition may favor one user group over another [Strong and Volkoff 2005]. Apparently, the difficulties of meeting the needs of diverse user groups with a single system [Markus 1983; Strong and Miller 1995] are further exacerbated with ERP [Strong and Volkoff 2005] and certain trade-offs between user groups have to be made.

Previous research has also recognized the need and developed methods for making trade-offs between data quality dimensions such as accuracy versus timeliness [Ballou and Pazer 1995] and completeness versus consistency [Ballou and Pazer 2003]. However, trade-offs are usually made within an ERP system without using these principled methods. For example, certain users are found to withhold data entry as long as possible because they are afraid of the adverse effects to downstream users and difficulties in cleaning up inaccurate data afterwards [Strong and Volkoff 2005]. As a result, timeliness is reduced despite the desire and expectation of using ERP to improve timeliness of information.

A case study of a division within a large manufacturer in the U.S. finds that many reports in ERP are difficult to develop and are error ridden [Vosburg and Kumar 2001]. The main reason for this problem is that the data required for reports are buried in a large number database tables with similar attributes, making them very difficult to identify. Another case study of two Australian firms in transportation and mining industries [Xu et al. 2002] finds that the use of ERP systems can introduce “new” data quality problems, such as reduced information accessibility due to tightened security and reduced data relevancy. It also finds that the increased complexity and lack of flexibility within ERP may cause data quality problems especially when the users are not trained on how to correctly handle exceptions. A similar problem is reported [Gattiker and Goodhue 2004] where issues in cost calculations remain unresolved because no one is certain whether making changes in one place would adversely affect users in other places.

In addition to increased complexity, ERP systems also make work processes within an organization more tightly coupled. Tight coupling means that there is no slack or buffer given between two items [Perrow 1999]. The tight coupling effects of ERP systems are investigated [Volkoff et al. 2005]. The study identifies three types of tight coupling of processes, which correspond to the three types of organizational interdependence [Mintzberg 1979; Thompson 1967]: pooled, sequential, and reciprocal. *Pooled coupling* occurs with common ERP services shared by similar business units. *Sequential coupling* occurs across multiple stages of a business process where the output of one stage is the input of the next stage. *Reciprocal coupling* occurs across different functional areas with their outputs becoming inputs of each other. These forms of tight coupling have positive as well as negative effects on data quality. For example, pooled coupling is introduced by the integration of various materials planning groups that use assembly instructions and part lead times to automatically determine the parts to be ordered. This integration improves the usability of inventory data for the management, but it increases the propensity of errors in the design group because they need to make frequent design changes; if these changes are not updated in the systems, the ordering system shared by all material planning groups may end up ordering wrong parts.

In summary, data quality problems in ERP are complicated and need a careful investigation. This study is a response to the call for research on solving the backlog of data quality problems [Lee 2003]. As the slack time between major processes for resolving problems continues to shrink, we must understand the root causes of data quality problems in ERP and develop effective solutions.

2.3. Impact of Organizational Context

Prior research suggests that organizational context is a determinant of information systems success [Ein-Dor and Segev 1978; Raymond 1990; Schultz and Slevin 1975]. Many organizational factors such as size, maturity, resources, and time frame [Schultz and Slevin 1975] have been found important to the performance of information systems. Other factors such as strategy, structure, environment, technology, task, and individual characteristics are also related to IS success [Weil and Olson 1989]. In ERP research, the fit between the organizational context and ERP is found critical for ERP implementation success [Hong and Kim 2002] as well as business values of ERP [Gattiker and Goodhue 2005]. Organizational fit has been examined to identify influencing factors such as data, process, output [Soh et al. 2000], user interface [Hong and Kim 2002], and interdependence and differentiation among subunits of an organization [Gattiker and Goodhue 2005]. Although data quality is an important aspect of ERP performance, how organizational context impacts data quality is not studied. Our research seeks to identify the organizational factors and ERP characteristics that cause data quality problems.

2.4. Normal Accident Theory (NAT)

In NAT, an accident is defined as "an event that is unintended, unfortunate, damages people or objects, affects the functioning of the system of interests, and is nontrivial" [Perrow 1999]. Normal accident theory holds that accidents are inevitable in complex, tightly coupled technological systems.

Complex interaction. A system is complex if its components have interactions of "unfamiliar sequences, or unplanned and unexpected sequences, and either not visible or not immediately comprehensible." Complex interaction is caused by many factors. The characteristics of complex interactions that are more relevant to data process in ERP-enabled manufacturing systems include [Perrow 1999]: unfamiliar or unintended feedback loops, many control parameters with unexpected interactions, indirect or inferential information sources, and limited understanding of certain processes.

Tight coupling. Tight coupling is especially likely in systems with processes that are unifinal (only one method to achieve a goal), invariant (where the order of sequence cannot be changed), and time dependent. Characteristics of tight coupling systems include [Perrow 1999]: time-dependent processes (no delay), invariant sequences (sequence), only one way/process to reach the production goal (unifinality), and little slack (criticality).

Complexity can cause unexpected and unobserved interactions between independent failures. These initial interactions can escalate into a system breakdown sooner or later in a tightly coupled system. The combination of complexity and tight coupling tends to make accidents inevitable.

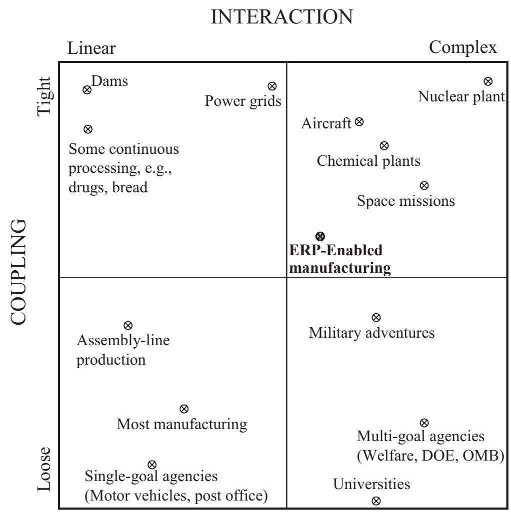

Fig. 1. Interaction/Coupling chart (Adapted from Perrow [1999]).

Interaction/coupling chart. Systems can be categorized along the *interaction* and *coupling* dimensions and put into an Interaction/Coupling (I/C) chart. An I/C chart adapted from Perrow [1999] is shown in Figure 1. Following the broader view of complex adaptive systems [Jacucci et al. 2006], a system includes not only the tangible technical components but also the organizational and social contexts within which the technical components operate. With this view, even organizations (such as universities) are considered systems. The upper-right quadrant of Figure 1 contains several examples of tightly coupled systems with complex interactions (e.g., space missions and nuclear power plants). In Perrow [1999], manufacturing systems are placed in the quadrant of loose coupling and linear interaction. This is probably true when manufacturers use separate information systems for different organizational units and functional areas. With the use of ERP systems, organizations, especially manufacturers, tend to move to the tight coupling [Volkoff et al. 2005] and complex interaction quadrant. Our case study will show that this is the case and being in this quadrant makes it prone to having certain data quality problems.

Application of NAT. Despite the criticism of having imprecise definitions and lacking criteria for quantifying complexity [Hopkins 1999], NAT has influenced research in many important areas. Some examples of the studies include the sociological study of risk analysis [Short 1984], the management of risky technology at NASA [Vaughan 1996], the organizational production and management of environmental hazard [Clarke 1989], the safe management of nuclear weapons [Sagan 1993], sense making and control in organizations with complex technology [Weick 1995], the nature of technology [Weick 1990], and the management of industrial crises [Weick 2001]. By placing a meaningful frame around flows of events, NAT allows us to analyze the complexity of technological organizations in the presence of the unexpected and helps us link multiple levels of analysis to better understand the issues involved [Weick 2004]. Coupling and complexity are key dimensions of organization diversity and modes of operation. NAT shifts the focus of explanation of accident to structural factors and combinations of problems instead of on isolated errors of individual human operators or design flaws in individual components.

Computer-related accidents have caused harm to the environment, injuries, and fatalities [MacKenzie 1994; Neumann 1995; Sammarco 2003]. It has been realized that the increasing dependence on software systems is creating the potential for loss of data or incorrect data. These data quality issues were considered as a new type of hazard that can lead to unacceptable physical, scientific, or financial losses [Leveson 2004]. For example, an incorrect data in airplane attitude control software led to the issuance of an incorrect control command, which caused an accident. The potential of using NAT to analyze software was recognized in Perrow [1999]. However, it is not until recently that NAT has been used to analyze the risks and complexity of software and information systems. For example, Lally [2002, 2005] developed conceptual models to explain the risks of information technology using NAT. She [Lally 2002, 2005] argues that frequent rapid changes in both IT systems and the work processes supported by the IT systems can further exacerbate the potential for disaster. NAT is also used in software design to better manage software complexity and coupling [Belotsky and Johns 2008].

We are not aware of any existing research that uses NAT to analyze ERP-related data quality issues. The limited existing literature on data quality in ERP systems seems to suggest that certain data quality problems can be introduced by the ERP systems because of their complexity and their tight coupling effects on the organizations and the processes [Volkoff et al. 2005; Xu et al. 2002]. However, the relationships among the organizational context, the characteristics of ERP systems, and data quality are not well understood. We address this gap in this research. To elucidate this relationship, we extend NAT in the context of accident and risk analysis to data quality management. We use this theory because data quality problems are similar to the notion of *accidents* used in Perrow [1999]. In addition, poor data quality can sometimes lead to disastrous accidents [Fisher and Kingma 2001].

In this research, we define *data quality accident* in ERP as a nontrivial data quality problem that damages the functioning of the ERP system and businesses processes. For example, delay or error in work orders may cause scheduling problems of the downstream processes. *Normal accidents* are data quality accidents that are inevitable and caused by the complex interaction and tight coupling of ERP within a certain organizational context. Next we report our case study that examines data quality problems in ERP using NAT as a theoretical lens.

3. RESEARCH METHOD

3.1. Research Design

We applied the case study research method to investigate data quality in an ERP system in a large manufacturing company headquartered in China. Since our research is exploratory rather than confirmatory, it is appropriate to use case study to delve into the complexity of data quality problems in ERP, and develop rich and informative conclusions in real-life context [Yin 2003]. Yin [2003] defines a case study as "an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident." Empirically investigating data quality problems in ERP systems aligns well with the preceding description. Case research is useful when "a phenomenon is broad and complex, where the existing body of knowledge is insufficient to permit the posing of causal questions, when a holistic, in-depth investigation is needed, and when a phenomenon cannot be studied outside the context in which it occurs" [Paré 2004], (page 4). Our study is exploratory in that it focuses on inductive theory building. The choice of a manufacturing company (given the pseudonym AC Company in this article) was based on the principle of theoretical sampling, which specifies that chosen cases should clearly represent the phenomenon under study in its natural social context [Mason 2002; Yin 2003]. In accordance with this principle, we judge AC Company to be an appropriate site for gathering retrospective data over 8 years of ERP operation. Since the initial ERP implementation in 2001, AC Company had gone through rapid expansion within a dynamic market environment. Data quality problems were reported and actions were taken to improve data quality reactively in an ad hoc fashion. The case setting thus provided an opportunity to explore problems in data quality and the factors that influence data quality.

Founded in 1991, AC Company has become the world's largest specialized air conditioner company with vertically integrated functions such as R&D, manufacturing, sales, and services. It was one of the "Top 100 Chinese Listed Companies" ranked by Fortune Magazine in the most recent eight years. AC Company has six branches in China, Vietnam, and Brazil, with more than 40,000 employees worldwide. The company focuses on new products development and innovation. Their slogan is "created-in-China instead of made-in-China."

Before ERP was implemented in 2001, the company had developed its own standalone software applications for the major functional areas such as accounting, purchasing management, after-sale service, and inventory control. Manufacturing was the only missing function. These applications were developed gradually over time. There was no enterprise architecture to support the fast growth of the company. In 1998, the company realized that the legacy systems were not able to support the growing business. Around that time, some of its competitors started to implement ERP. These factors led to the company's decision to adopt ERP. In 1999, AC Company initiated the ERP project. The system was successfully implemented and went live in 2001. One of the major challenges in implementation was a data problem: the configuration of ERP parameters. Most of these parameters did not exist before. For example, it took a long time in the logistics department to configure all the data elements related to materials issuing and receipts. The ERP system was upgraded in 2008 to accommodate the demand for system capacity and performance. The ERP project was considered a success by the company. Among many benefits harnessed from ERP, one was the standardization of data.

3.2. Data Collection

Data collection and analysis were conducted in an iterative fashion between December 2007 and August 2009 via interviews and the review of a large amount of internal documentation provided by AC Company. Twenty-four interviews were conducted with the top managers, the IT director, IT project managers, IT staff, an ERP consultant, department mangers, and system users in all functional departments (Finance and Accounting, Production Planning, Logistics, Manufacturing, and Marketing). The initial interviews were done in December 2007 with the IT director and IT developers and a few system users. From January to May 2008, two telephone interviews with the IT director were conducted. Documents were collected and reviewed. In May 2008, the first author visited the company again and interviewed more system users, the IT director, department managers, consultants, additional developers, and the top managers. From June 2008 to August 2009, more data were collected through phone calls, emails, and documents. Table I lists the roles of respondents and the number of interviews conducted with each respondent.

The interview protocol contained a series of open-ended questions to encourage discussion of topics relevant to data quality. Included were questions about respondents' perspective for using ERP, their past work and training practices, the data quality problems they experienced, and the actions they took to deal with the problems. Table II is an excerpt of the protocol. Each interview lasted approximately 90 minutes, on

Table I. Description of Sampling

Role	# of Respondents	# of Interviews per Respondents
Top managers		
CIO	1	1
CEO	1	1
IT director	1	5
IT project managers	2	1
IT staff	4	1
ERP consultant	1	2
Department mangers	4	1
System users	5	1
Total interviews		24

Table II. Excerpt of Interview Protocol

Who made the decision to have an ERP system? Was it from the top management or from the IT department?
Do you think the ERP is a success? Did the system meet your initial expectations?
What are the data quality problems in the system? How are they solved?
What are the problems in data production? How are they solved?
What are the problems in data storage and maintenance? How are they solved?
What are the problems in data utilization? How are they solved?

average. All the interviews were recorded and transcribed. Data collection was tightly interwoven with data analysis. Subsequent interviews were conducted to follow-up on the insights that emerged from the data analysis. New data were used to refine and modify prior concepts and guide subsequent data collection. Data collection and coding were done until "theoretical saturation" [Eisenhardt 1989] was reached, at which point additional data did not add to the concepts and subcategories.

3.3. Data Analysis

We analyzed the data iteratively, alternating data coding with investigation of theories that fit the emerging interpretation. We first translated transcripts of the audio taped interviews from Chinese into English. Interview data was coded to reflect constructs that were identified from our review of literature (data quality problems in data production, data storage and maintenance, and data unitization, organizational context, interactive complexity, and tight coupling of ERP systems). Both authors conferred on the meaning of statements made during interviews and compared interpretations with reference to the theory. The codes were then triangulated using the Company's documents such as transaction summary reports and management reports. Finally, all the codes and notes were carefully examined to identify a set of major themes [Patton 1990] which form the framework presented in Section 5. These steps provided multiple opportunities to reflect upon the data, generating further insights. Such a process helps ensure that the conceptual framework we have developed satisfies established criteria for the credibility and validity of qualitative research [Miles and Huberman 1994; Pawlowski and Robey 2004]. No attempt is made to statistically evaluate the strength of concepts included in the framework [Strauss and Corbin 1990]. Though the findings from such analyses are detailed and particularistic, they can be used to develop a general explanation of the results [Eisenhardt 1989; Orlikowski 1993] through analytic generalization [Yin 2003]. We presented our findings to key informants including the IT director and system users of the company. Based on their feedback, we made minor adjustments to our analysis.

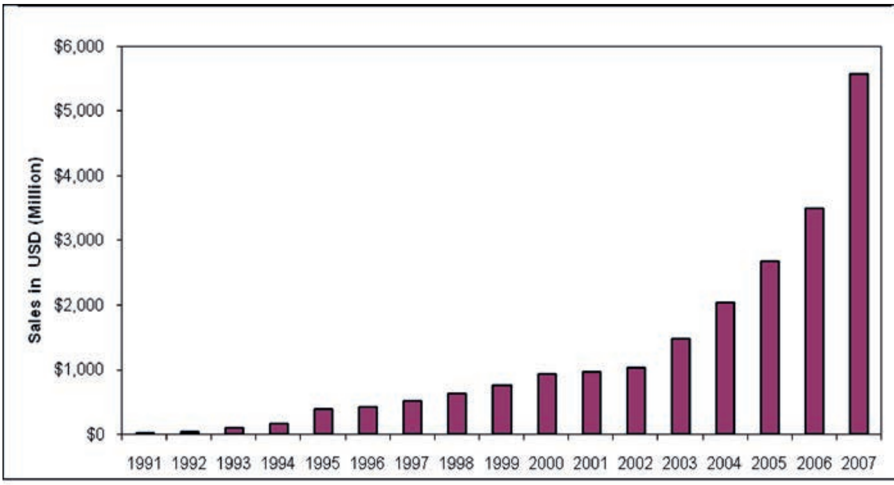

Fig. 2. Sales of AC Company.

4. RESULTS

Through our case study, we find that certain ERP characteristics within a specific organizational context can cause normal accidents in data quality.

4.1. Organizational Context

4.1.1. Rapid Growth. The ERP system had successfully coped with the faster growth after it was in operation in 2001 (Figure 2). The revenue of AC Company increased from one billion dollars in 2001 to 5.7 billion in 2007. The impact of fast growth on the system was twofold. On one hand, the fast growth resulted in excessive pressure on IT infrastructure. The ERP system had to be extended in both capacity and capability to support the growth. The amount of data produced and processed increased exponentially. The system had to be reconfigured in its hardware, software, and architecture design to cope with the growth. For example, the centralized data model was replaced with a distributed data model. On the other hand, the business processes have been evolving. The changing processes require frequent reconfiguration of the system. As a result, there has been complex interaction between business needs and the system enhancement.

4.1.2. Dynamic Business Environment. AC Company faced a very turbulent market caused by several factors such as government policies, major economic events (e.g., global economic crisis), other related industries such as real estate, marketing strategies of competitors, and even weather.

The fast-changing market required quick responses from almost all departments. For example, the planning department had to adjust work plans on a daily basis as described by the manager of planning: “ERP is strong at planning. . .if it works in a stable business environment. But for us, the market is changing at a fast pace. The sales department may have given us an order a month ahead of the delivery time, but they often change their mind and want it delivered in two days. He would say: ‘This is an urgent order, the market needs it immediately, you must work on this right away.’ Then we have to adjust working orders in ERP to deal with such abnormal situations.”

The engineering department also needed to make frequent engineering changes to cope with market pressure. There were many reasons for engineering changes such as design errors, customer requirement changes, product quality problems, and changes in competitor's products or materials. Specifically, to respond quickly to a dynamic market, new products were developed in short cycles, which caused certain design errors that led to engineering changes. As explained by the manager of R&D: "We have a very short development cycle, to respond to the market needs... We have more than 1,000 new models each year. The design of a new model may contain defects that will be reported through feedback from the market. These defects need to be fixed immediately through design changes. We cannot afford to wait for the next model to get it fixed." As a result, engineering changes happened during the whole manufacturing process, not only within design phases.

The dynamic business environment also caused frequent changes in business processes and the ERP system. These changes increased the potential for an incident to propagate and grow into an accident. Furthermore, the dynamic business environment added complexity to the ERP system in hope to respond to rapid changes. For example, new components were added to the system to handle changes of product orders at any manufacturing stage. However, such ad hoc functionality was often implemented without a robust design. As a result, it was usually buggy, causing unintended adverse consequences.

4.1.3. Market Share Focused. At AC Company, the organizational goal was to increase market share. As described by the CEO: "Our market share is about 40% and we are striving for higher. There are 20–30 air conditioner companies around ...The competition is intense and we have to do everything to keep our leading position and grow."

As a result, meeting customer and market needs was the first priority. As described by the marketing manger: "The competition is so tough, we cannot afford to tell our customers to wait for two weeks when they came to us. We have to deliver whatever the market demands." To meet the ever-changing market and customer requirements, operations were often interrupted with exceptions and all the manufacturing activities had to be adjusted frequently. Insufficient attention and resources were allocated to improve the operation performance, as explained by the IT director: "Our goal [of increasing market share] determines the management strategies. Currently we are focusing more on sales and product development than on internal operations. It is an issue of investment of resources."

Among other factors, cost reduction was perceived as very important to stay competitive. As a result, new materials were constantly being adopted and suppliers were switched frequently. Frequent changes in materials and suppliers caused frequent design changes and product quality issues. This strategy resulted in frequent abnormal situations of the system that might trigger normal accidents. Moreover, the market-share-focused strategy led to limited resources available for improving the system performance. For example, many users had limited understanding on the system due to insufficient training. The system was never thoroughly tuned to improve its performance.

4.2. ERP Characteristics

4.2.1. Interactive Complexity. First, there are unintended interactions of data in an ERP system. For example, material plans were generated automatically based on BOM and MPS (Master Production Plans). With material plans, purchase orders were generated automatically according to predefined inventory control policies. When there was a change in BOM, the material plans had to be updated accordingly. If BOM changes during the planning phase, the purchase order would be modified to remove the unneeded materials and add new materials. The interactions among BOM, MPS, and inventory control policies were well understood in AC Company. However, when BOM changed during the purchasing and assembly phases, the purchase orders had been released to the vendors, and sometimes the materials had already been purchased. This caused certain materials to be wasted because they were not needed anymore. On the other hand, the required materials were not available when they were needed and the production process for that product was interrupted. The interaction between BOM and purchasing orders during the purchasing and assembly phases was not intended in ERP.

Second, ERP contains many control variables with potential interactions. Control data in ERP came from multiple sources. For example, there were 390,000 material items used at AC Company. The data came from and were used by several business units such as design, purchasing, sales, manufacturing, costing, and planning. As a result, these variables may have unexpected interactions because any change in one material item can affect several business units.

Third, certain data came from indirect or inferential information sources in ERP. As explained by the manager of planning: "We are relying on ERP to get the production information, material demand information, and planning information. . .there are always new problems that cannot be solved within the current business process. ERP is tightly bundled with business process; operations are all interconnected. As a result, many times when a user encounters an abnormal situation, he has difficulty to evaluate the impact of the actions he is about to take; he knows that certain operations would be affected but he doesn't know what the impacts would be."

Finally, after years of learning, there was still limited understanding of certain data processes in ERP software itself. The ERP system contained a large number of modules and sophisticated data structures. The source code was not available to AC Company. IT personnel learned the data structure of the ERP system over time. In addition, IT personnel needed to have a thorough understanding of the business processes to support the operation of the system. However, the learning took a long time and was not guaranteed. For example, after the ERP system was upgraded to the next version in 2008, the accounting department found that a report contained different data when it was displayed in another format. Through further testing, they found out that the code generating the report in Excel format missed certain data elements. Similar problems often occurred in customized functions developed internally as well as some of the standard functions that came with the software upgrade. The data structure of the new version was totally different from that of the previous version; IT personnel had limited understanding of it even after six months of development and testing. As the accounting manager explained: "during the customization process, the development team was exploring the system while working on it at the same time. So maybe our developers and support [personnel] were not very clear [about the system]. There were some problems with the correctness of the data in our reports. Also sometimes they would not customize the system because they were not sure about the consequences."

4.2.2. Tight Coupling. ERP integrates the traditional stand-alone systems and considers them to be a part of interlinked processes that make up the business. There are interwoven interactions among system components, business units, and subprocesses. As a result, ERP-enabled manufacturing is largely a tightly coupled system.

First, data and business processes in ERP-enabled manufacturing were more time dependent. As noted by the manager of logistics: "ERP requires coordination amongst the departments... The timeliness and accuracy of data are very important. Before adopting ERP, it did not matter if you completed your work early or late. It is different now. If you do not complete on time, downstream people cannot start their work."

Second, the sequences in the data process of the ERP system were invariant and unifinal. The manager of the finance department commented: "Management must follow the workflow implemented in the system. . . . We must have a production plan, followed by a purchasing and acquisition plan, product quality inspection, products checked into storage, and finally the accounting of material consumed. Management is integrated with these interlinked processes, which are controlled by the system, not people." Apparently, people at AC Company were aware of sequential coupling introduced by ERP.

Finally, there was little slack in ERP-enabled manufacturing. As a result, data must be on time and accurate, as the manager of logistics commented: "Before ERP, we [different departments] were working independently... but ERP is an integrated system which requires a unified standard. For example, it was not a big deal if the data was entered later [before ERP], but now things are different. The downstream departments cannot work if I don't enter the data on time."

In summary, ERP-enabled manufacturing is a complex and tightly coupled system. Within such systems, certain data quality problems appear inevitable.

4.3. Data Quality Problems at AC Company

The organizational context and ERP characteristics have caused data quality problems at AC Company. Data was frequently reported as delayed or incorrect. Data quality has huge impacts on the effectiveness of the ERP system. Our study identified several data quality problems in the ERP system at AC Company.

Three work roles in data processing have been identified: data *c*ollector, data *c*ustodian, and data *c*onsumer, which are called the 3C's of data [Lee and Strong 2003]. *Data collectors* provide initial input of organizational data; *data custodians* are responsible for storage and maintenance of the data; and *data consumers* utilize the data for further integration, aggregation, presentation, and interpretation of data. The quality of the same data is often perceived differently by people in different roles [Lee et al. 2002]. The three work roles are involved in three data processes: data production, data storage and maintenance, and data utilization. In this section we discuss data quality problems in the three data processes.

4.3.1. Data Production. Data production involved collecting and entering data into the system. There were two types of data: transaction data and control data. *Transaction data* included data about business activities such as customer orders, work orders, transfers between onsite storages of intermediate products, invoices, etc. The data sources of transaction data were simple but the volume was huge. *Control data* included system parameters and primary data for materials, BOM, item routines, and materials planning parameters. Although the amount of control data was relatively small, it was complex because it involved multiple data sources and many operations and business processes depended on the control data.

Transaction data quality problems: Timeliness, inaccuracy, amount of data. At AC Company, the biggest problem for transaction data was timeliness and amount of data. Lack of timeliness also caused inaccuracy in data because of sequential coupling introduced by the ERP system. Delay was reported in several data production points. To illustrate the problem, we present a simplified workflow of the manufacturing process in Figure 3. The process had roughly four subprocesses (shown as rectangles), two warehouses (shown as rounded squares), and four types of inventory storages (shown as ovals). Materials were purchased and stored in a raw material warehouse. Based on the work orders, they were distributed to the work-in-process storages (which were located onsite at manufacturing shop floors). To record this type of transactions, the system deducted the inventory in the raw material warehouse and increased the inventory in work-in-process storages. After the manufacturing sub-process, work orders were recorded as completed in the system. Accordingly, the system deducted the inventory

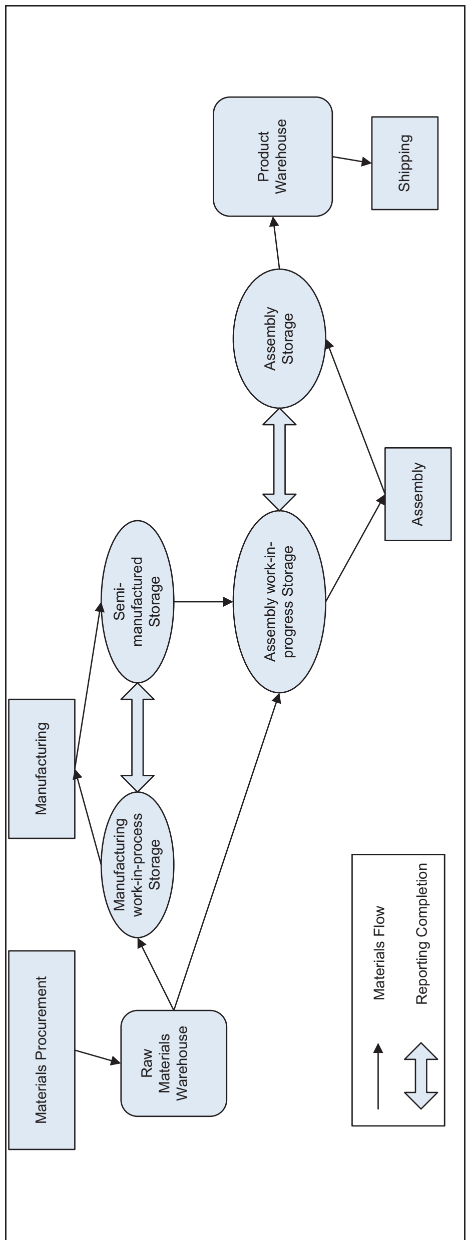

Fig. 3. Overview of workflow.

Table III. Example of Completion Report Time (2-11-2001)

	Shop Floor 1		Shop Floor 2		Shop Floor 3		Shop Floor 4		Assembly	
	Day	Night	Day	Night	Day	Night	Day	Night	Day	Night
Monday	nd* 9:47	9:50	18:24	**delay	18:59	8:55	delay	8:48	19:23	8:41
Tuesday	22:10	10:10	nd 12:19	12:19	20:01	8:53	20:24	8:55	20:29	8:39
Wednesday	21:36	8:24	nd 10:51	10:51	21:34	9:08	19:49	9:06	20:25	9:11
Thursday	21:10	9:24	nd 10:05	10:05	20:35	9:01	20:06	9:33	21:00	9:03
Friday	21:52	9:28	nd 10:40	10:40	20:35	9:00	20:04	9:25	20:46	9:04

*nd – next day. **delay – completion not reported at the report time.

in work-in-process storage and increased the inventory in a semimanufactured warehouse. A similar procedure happened for the assembly subprocess.

One type of delay in data production was the delay in recording completion of work orders. This delay was reported to be a major problem since it affected downstream processes. Completion reporting was done in batch mode in the offices by the supervisors several times each shift. It was not done online, onsite, as the products were finished and taken off of the production lines. Therefore, the on-time completion reporting depended on factors such as managerial and operational policies as well abnormal situations in the manufacturing process that prevented the supervisors from entering the data on time.

Table III is an example of a weekly report of completion report time three months after ERP was implemented. Day shifts should report completion before 9pm and night shifts should report before 9:30am. Delays are marked with a shaded background in the table.

We can see that delays occurred at shop floors 1 and 2 almost everyday. The documented reason was the absence of system users. However, the IT director told us: "I think it was because the users were not very familiar with the system and also the responsibility of reporting was not stabilized. There might be other reasons unreported such as insufficient training, undefined operation procedures, etc."

Delay of completion reporting had several consequences. After the products were taken off of the production line, materials had been distributed and consumed. However, the system did not show this information without reporting completion of work orders. The inventory in work-in-process storage was not deducted even though the materials had been consumed. The inventory in semimanufactured storage was not increased even though they had been produced. This caused problems for downstream processes. For example, the semimanufactured products needed in other work orders could not be scheduled if these products were shown as unavailable in the system.

Another type of delay in data production was delay in transaction receipts processing, which caused mismatch between records in the system and reality, a form of data accuracy problem. For example, a material had been distributed from the raw material warehouse to shop floors and it had already been consumed in production, however, this transaction was not entered into the system on time. As a result, work-in-process storage had no record of receiving this material. When the system deducted the inventory in work-in-process storage during the completion reporting, it could not find this material in work-in-process storage. This was called "unsuccessful back flushing", which is a failed retrospective update caused by inconsistent data. Retrospective update was a transaction in the system to indicate the material had been consumed (according to BOM) when a product was added to the finished product storage. Table IV provides an example of the number of unsuccessful retrospective updates for a specific day.

There were two factors that led to delay and accuracy problems: (1) the amount of data, and (2) the paper-based data production process. For example, there were about

Table IV. Example of Number of Unsuccessful Retrospective Updates (2-11-2001)

Shop Floor 1	Shop Floor 2	Shop Floor 3	Shop Floor 4	Shop Floor 5	Assembly
6	35	55	40	34	98

100,000 transactions every day in the raw material warehouse. These transactions data were all paper based. The warehouse operators dispatched the materials based on the checkout lists. These checkout lists were then summarized by the supervisors based on the distribution lists generated from the system. A material item in the distribution lists might have several transactions. The supervisors summed up the transactions and added a total number. Then the summarized transaction data (still on paper) was passed to the account supervisors who were responsible for confirming the transactions and inputting the summarized transaction data into the system. One transaction was processed by three groups of people manually before it was entered into the system. Transaction data were often not entered into the system the same day. Supervisors sometimes made mistakes in their manual processing of transaction data.

To solve these data quality problems, AC Company took several actions. The CIO told us: "We have tried different things but none of them worked. We tried to strengthen the management, train our operators, etc. It worked for a while but as the transactions volume kept going up, it went back to the old way." Automating data production and the collection process would eliminate the need of manual data entry and improve data timeliness and accuracy. However, over the past eight years of ERP operation, the data collection was not automated. One major reason was that data automation cost was not justified as cost efficient by the management. The IT manager summarized: "Theoretically it is not much to discuss about data quality [in data production]; the point is, if we would like to accept the cost of data production."

Control data quality problems: Inaccuracy, questionable believability, and volatility. Different from transaction data, the biggest data quality problems for control data were accuracy, believability, and stability. Control data came from multiple data sources. For example, material items data came from product design, purchasing, sales, manufacturing, cost analysis, and planning. Adjustment in any of the data sources needed to be updated in the system. The accuracy and believability of control data depended on the data production process, operators' knowledge about the system and business processes, and communication among different business units. Control data were also used in many other activities. Thus any quality problem in control data would have broad adverse effects due to pooled and reciprocal coupling introduced by the ERP system.

Here we use BOM as an example to illustrate the problem. At AC Company, BOM and material items data were highly volatile. Their volatility caused problems to the whole manufacturing process. AC Company had to react quickly to frequent engineering changes. This was the nature of the industry in a turbulent business environment where product innovation is a competitive necessity, production operations are constantly being reconfigured, and response time is critical. Engineering changes required BOM to be changed accordingly. MRP II (Manufacturing Resource Planning), the core module of ERP, was based on BOM and other information such as demand forecasts to calculate required materials and generate product orders, work orders, purchase orders, etc. All of these data had to be adjusted if BOM changed. However, frequent BOM changes made this adjustment very difficult. In addition, the adjustment caused inaccurate data in almost all the processes such as production planning, logistics, manufacturing and cost analysis, and pricing. The manager of production planning commented: "We make design changes whenever the customer requested. It is easy to inject errors in making the changes. The BOM changes constantly, from product development until it is retired from the market. This has huge impacts on the [ERP] system. This is the weakest link. It doesn't matter how hard we work. If the BOM changes, the base [production plans] that ERP relies on to generate materials requests, work orders, as well as all kinds of financial analysis has to change frequently." Frequent engineering changes also injected errors in BOM when the changes were not reflected correctly or on time in BOM. Another consequence of the volatility and inaccuracy of BOM was the excess materials, namely those that were removed from the BOM after it was purchased. This was another "weakest link" according to the director of planning.

To solve this problem, AC Company set up a 30-person group to maintain BOM and material items data, which changed on a daily basis. The group was responsible for bridging design and manufacturing by transferring design details into BOM and material items data. There were more than 60 million data points for material items (142 properties each material, about 400,000 different materials) and 120 million data points for BOM (20 properties per record, more than 6 million records). However, the frequent and late-stage engineering changes were still an unsolved problem, as noted by the director of planning: "There is nothing we can do; we have to follow it (the BOM change). Has it changed today? How about tomorrow?" Even the CIO admitted that there was not an easy solution to this problem. "There is a conflict between the (ERP) system and the managerial strategy. ERP software requires a stable manufacturing environment but our operation needs a system that is really flexible. . .We just have to accept it [quality problems in control data]."

4.3.2. Data Quality Problems in Data Storage and Maintenance: Inconsistent, Incomplete and Incorrect Data. Data quality problems in data storage and maintenance were also observed at AC Company. The IT department was responsible for data storage and maintenance. The major problems were data completeness, consistency, and correctness. These problems were mainly related to the complicated data structure and data processing logic of the system. Ideally a stable and reliable ERP system should have little problem in data processing. However, the specific organizational context at AC Company required the ERP system to be easily configurable to support its dynamic processes. An ERP system provides this configurability by its phenomenally large data model and complicated data structure. For example, at AC Company, a simple transaction such as shipping of an order involved multiple tables in work orders, sales, finance, and planning in the database. The transaction processing capability of the database component of the ERP had the capability of handling such transactions to ensure the data integrity. However, automatic rollback sometimes could fail and required manual processing. In addition, interruptions to transaction processing sometimes occurred when new configurations and new programs were deployed. Manual interventions of failed transactions were often mishandled, as explained by the IT director: "The system has a rebuild function that can be activated to check and repair (corrupt data) automatically; the software vendor expected this might happen. However, the users might not take it seriously to do anything about it when it happens, it is not a common problem." On the surface, it was a user problem, but the problem came from unfamiliar interactions within the system. It happened infrequently and most of the users had limited experience or knowledge about it.

Another problem we observed at AC Company was the design and programming errors during ERP customization. One example of this problem was the developer used "order total amount" instead of "actual sales amount" in developing a customized sales report. This error was not identified in testing and was captured in operation. The IT department enhanced quality control by adding user acceptance tests in the development process. Although such measures had reduced programming errors, they could not completely eliminate errors.

4.3.3. Data Quality Problems in Data Utilization: Little Value-Added. Data utilization is about realizing the value of data by making decisions and taking actions based on it. Little value-added was a problem observed in data utilization. Two factors that determined the usefulness of the data were: (1) the quality of transaction and control data, and (2) the quality of data models (such as forecasting models).

Instead of the scheduling module of the ERP system, the planning department used a spreadsheet application developed by the IT department to generate production and purchasing orders. The scheduling module was not usable because of various data quality problems in the transaction and control data discussed earlier. For example, volatility had caused the "order time" parameter (time when an order was entered) in the ERP system to be unusable. In the spreadsheet application, the planning department had to create their own order time parameters for production planning and material purchasing.

In the accounting department, cost analysis and pricing relied on data from all the upstream departments. The lack of accuracy and believability in materials and planning data combined with the company's unique manufacturing processes caused the data generated from the system to offer little value-added to costing and pricing analysis. As explained by an accountant: "[The problem] is neither software nor people. I think it is the characteristic of our company. New product models come out all the time, but we have limited resources... sometimes (production) starts without knowing the total quantity to be produced, and we have to start pricing. As a result, the price generated by the system couldn't be used and we had to proceed manually. We had to look at design documents and estimated the cost of each item."

Another data quality problem in utilization was with sales forecast. ERP relied on sales forecast for long-term production planning. However, the sales forecasts were not accurate and no effort was put into improving the analytical model for forecasting. Frequent adjustments had to be made to the production plans in the system.

In summary, data quality problems existed during data production, storage and maintenance, and utilization. These problems seriously impacted the system and business effectiveness. Efforts had been made to solve these problems without a satisfying result. Next we discuss if these data quality problems are normal accidents and how they can be mitigated.

5. DISCUSSION

We investigate the research questions presented in the Introduction by identifying the data quality problems and factors that contribute to the problems. Based on normal accident theory, accidents are inevitable in complex, tightly coupled technological systems. It is important to examine if the preceding identified data quality problems are normal accidents in order to address them. In this section, we first discuss the organizational factors and ERP characteristics that cause each data quality problem. We focus on finding the root causes of these problems to decide if a problem is a normal accident or a problem that can be avoided. Then we propose two strategies to address these problems.

5.1. Normal Accidents of Data Quality Problems

Data quality problems in data production. Data quality problems in data production were caused by the organizational context that involved mismatch between task and technology [Goodhue and Thompson 1995]. The timeliness and amount of data problems of transaction data were caused by unmatched data volume and data processing mechanisms. The problems might be solved by automating the data production process. Currently, AC Company is implementing a Manufacturing Execution System (MES). Although the main purpose of the MES is not for data collection, it can be used to collect production data automatically. MES has been piloted in one assembly line and results showed that it largely improved data timeliness and accuracy. AC Company had tried different approaches over the past eight years to solve these problems but none was successful. They have learned from the experience that automating data production is the only solution. However, it took the management a long time to justify the investment in data automation.

The inaccuracy, questionable believability, and volatility of control data were caused by the mismatch between frequent engineering changes determined by the organizational context and the stable working environment required by ERP. The management was focusing on current opportunities to expand the market at the price of compromised data quality and inefficient operations. This problem cannot be easily solved or avoided in the current organizational context. AC Company had taken actions to mitigate the problem but soon found that none of them helped.

While the data quality problems in data production were rooted in the mismatch between task and technology, the complex interactions and tight coupling among different business units, operators, and processes involved in an ERP system make them from independent "mistakes" to normal accidents. The inaccuracy problem of transaction data was caused by the sequential coupling of the system. Similarly, data quality problems in control data were caused by the pooled and reciprocal coupling of the system. For example, a change in BOM may propagate to unavoidable mistakes in purchase order, inventory data, as well as product orders.

Data quality problems in data storage and maintenance. On the surface, the inconsistent, incomplete, *and* incorrect data problems were caused by operators who did not take actions to reconcile corrupted data from failed database transactions. But in reality, these problems largely resulted from the organizational factors with their adverse effects amplified by the feedback loops of complex interactions between business and the ERP system. The rapid growth and dynamic business environment put excessive pressure on IT infrastructure, hardware, software, and architecture design. In addition, continuous changes in business processes had to be supported by continuous error-prone customization of the system.

These problems can be mitigated by several ways such as following a well-defined development process, rigorous testing, and training of developers. However, software development for a complex and tightly coupled system is essentially difficult [Brooks 1987].

The complex interaction and tight coupling of ERP system made these problems normal accidents. For example, the limited understanding of the complex system design prevented operators from learning about the consequences of not taking actions. As a result, these problems are general for all ERP systems in large organizations.

Data quality problems in data utilization. The problem of little value-added in data utilization was inherited from the data quality problems in transaction data and control data. Pooled and reciprocal coupling of the system caused this problem. At AC Company, much of the data aggregated by the system couldn't be directly used in guiding business decisions. A large amount of effort had to be spent on evaluating the correctness of the data.

Data quality problems in data utilization were normal accidents so they were not easy to solve with the organizational context. The dynamic business environment required frequent adjustments of data and processes, and the market-share-focused approach and rapid growth prevented the company from effectively using the data to develop usable analytical models (such as forecasting models).

In summary, the data problems identified in this research resulted from complex interaction and tight coupling of the ERP system as well as the specific organizational

Table V. Data Quality Problems

Data process		Problems	Organizational context	ERP characteristics
Data production	Transaction data	Timeliness, Inaccuracy, Amount of data	Rapid growth caused mismatch between volume of data and manual data process. Rapid growth resulted in complex interaction between business needs and system enhancement.	Complex interactions and tight coupling among different business units, operators and processes turn independent 'mistakes' into normal accidents.
	Control data	Inaccuracy, Questionable believability, Volatility	Dynamic business environment caused mismatch between frequent engineering changes and the stable working environment required by ERP. Dynamic business environment added complexity to the ERP system to respond to rapid changes.	
Data storage and maintenance		Inconsistent, Incomplete, Incorrect data	Rapid growth and dynamic business environment led to continuous customization of system to support new business process. Rapid growth resulted in complex interaction between business needs and system enhancement.	Complex interactions in ERP result in unintended and unfamiliar feedback loops that caused operators' limited understanding of the system.
Data utilization		Little value-added	Market share focused strategy prevented sufficient effort on improving quality of analytic data models. Market share focused strategy resulted in frequent abnormal situations of the system that might trigger normal accidents.	The normal accidents in transaction and control data and pooled and reciprocal coupling of feedback loops caused problems in data utilization.

context. These data quality problems are considered normal accidents. Table V summarizes our findings discussed earlier.

5.2. Strategies to Address Data Quality Problems

The previous analysis shows that certain data quality problems in an ERP-enabled manufacturing are normal accidents. This means that the company has to rethink solutions for data quality problems. NAT highlights factors that contribute to data quality problems even though it does not provide explicit solutions. The first and most critical step towards strategies to address these problems is to identify the root causes of the problems. This strategy aligns with the results in Lee [2003], which suggests that data quality problem solving is complex and should involve revising the rules embedded in the data structure and business process instead of focusing on fixing the

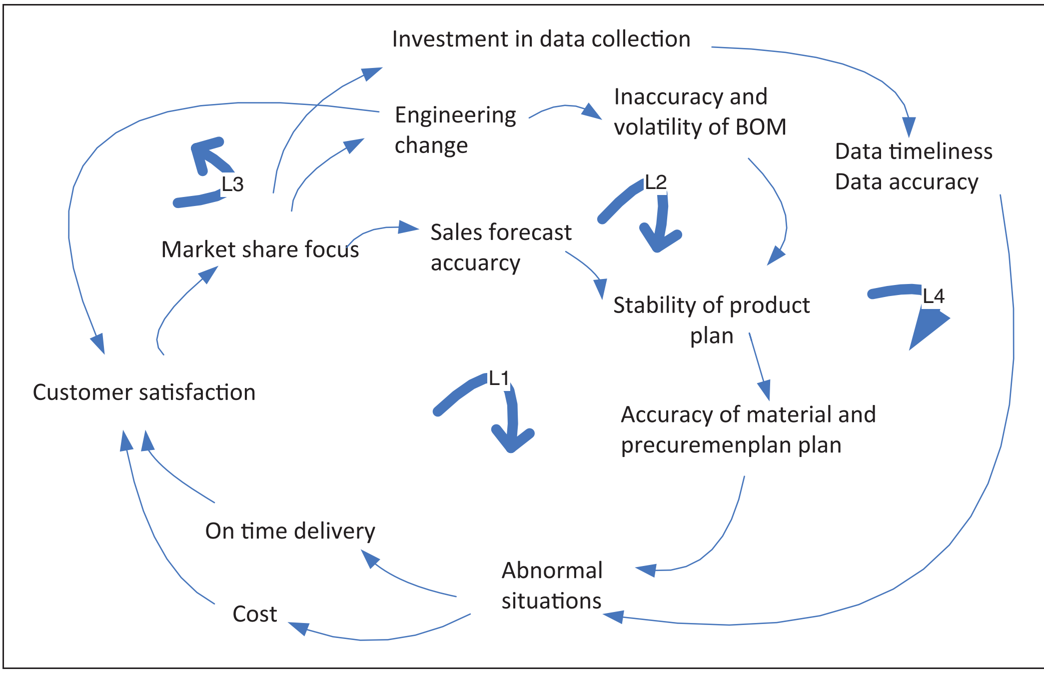

Fig. 4. Feedback loops of data quality.

isolated problems. However, it is usually hard to accomplish this as the rules are not recorded, shared, or understood.

To address these data quality problems, we propose two strategies to mitigate the impact of the two root causes, namely ERP characteristics and organizational context: (1) identifying hidden interactions to reduce the complexity of ERP, and (2) improving organizational fit of ERP to improve the organizational capability of reducing accidents.

Identifying hidden interactions. This strategy aims at identifying hidden interactions among major factors that have impacts on data quality problems. The purpose of the strategy is to reduce the complexity of the system by reducing the number of unexpected interactions. ERP is an integrated system that introduces many unfamiliar feedback loops and some of its processes are not well understood. We use the case study as an example to demonstrate how organizations can use this strategy. Figure 4 shows several such nonobvious feedback loops that affected data quality.

In Figure 4, Loop L1 shows the impact of inaccurate sales forecasts on the system. The organizational goal was to increase market share by unconditionally satisfying customer needs. Sales forecasting was not the focus of the sales department. Also the dynamic business environment made forecasting difficult. As a result, the sales forecasts accuracy was usually low. Stability of product plan is low as it had to be adjusted frequently to accommodate emerging situations. Accordingly, accuracy of materials and procurement plan was decreased because they had to be updated. This caused abnormal situations such as exceptions, unavailable materials, and insufficient capacity. The abnormal situations disturbed the whole manufacturing process, which in turn increased the cost and decreased the on-time delivery rate of customer orders. Customer satisfaction was hurt and in the long run, the market share would decrease. Organizations aware of the impacts of sales forecasts on the system can develop effective solutions to produce positive outcomes. For example, after installing the ERP system, a chemical manufacturer created a demand manager position responsible for demand forecast, production scheduling, and delivery estimation. The work of the demand manager has helped reducing cost and increasing customer satisfaction [Davenport 1998].

Loops L2 and L3 together show the impact of engineering changes on the system. The organizational goal of increasing market share led to meeting the market demands as the first priority. This led to frequent engineering changes to increase the customer satisfaction (L3). The engineering changes increased inaccuracy and volatility of BOM. Problems in BOM decreased the stability of the production plan and finally hindered the company from reaching its goals (L2). However, this loop (L2) was not visible to the management while Loop L3 was emphasized by the management as an approach to increase market share. The CEO told us: “We have to accept this [BOM changes]. We have to satisfy customers' requirements. It is a tough market; the existing models have to be updated all the time . . . with 30% sales increase annually, this is what we have to live with, we will do whatever it takes to reach our goal.”

Loop L4 shows the impact of investment in data production on the system that was also hidden from the management. The management was reluctant to invest in network, hardware, and software to automate data collection. As discussed in the previous section, this caused data timeliness *and* accuracy problems, which led to abnormal situations. However, the management justified their decision on paper-based data collection: “It [paper based] is the [least expensive way to collect data]. ERP does not require on-time process data, as long as we are ‘not late’ why should we spend half million or more (RMB) on computers?”

While each of the feedback loops is already difficult to observe, the interactions of these feedback loops make data quality problems much more difficult to comprehend. By explicitly identifying the hidden interactions as shown in Figure 4, management can make more rational decisions in addressing data quality problems.

Improving organizational fit of ERP. The purpose of the strategy is to reduce the mismatch between task and technology in an organizational context, which causes data quality problems. The strategy aims at improving organizational fit of ERP by changing the endogenous factors of organizational context such as management policies. Research has shown that organizational fit of ERP is critical for ERP implementation success [Hong and Kim 2002]. There are several actions that can be taken to improve the organizational fit of ERP to reduce the normal accidents of data quality.

First, people in different work roles of data process, for example, data collectors, data custodians, and data consumers, need to obtain both comprehensive understanding of the critical organizational processes and detailed knowledge of this very complex ERP software [Soh et al. 2000]. Research has identified three knowledge modes in solving data quality problems: *knowing-what* refers to the understanding of the activities involved in data production processes, *knowing-how* refers to the understanding of the procedures to handle known data quality problems, and *knowing-why* is the ability to analyze underlying principles and discover previously unknown data quality problems or solutions [Lee and Strong 2003]. Organizations should facilitate the knowledge acquisition process by encouraging users to learn about the right knowledge of the system and organizational processes.

Second, an organization must balance flexibility and operational performance. The market-share-focused strategy that emphasized flexibility and quick response to customer needs caused frequent changes of processes and data. Change has been recognized as one of the causes of normal accidents in information systems [Lally 2002, 2005]. These normal accidents will have adverse effects on the company in the long run. In fact, AC Company has been experiencing increasing complaints from customers and many complaints can be traced back to certain data quality problems.

Thus organizations should have a strategy that balances the flexibility and operational performance by having a controlled change process. A change should be analyzed on its criticality and impacts on the system before it is implemented. Another action to improve operational performance is to replace an outmoded process with an efficient one. For example, the manual data process of transaction data should be replaced by an automatic data collection process. Often the inhibitor is the managerial reluctance to provide resources for such a migration. This happens because the real cost of not migrating to the new process or system is not evaluated [Venkatraman 1994].

In summary, our research has identified major data quality problems in ERP-enabled manufacturing systems and the causes of the problems. We have also proposed strategies to mitigate these problems by changing certain aspects of the organizational context and coping with ERP characteristics. The findings are summarized in a framework (Figure 5). Certain data quality problems will be with us for a long time not only because organizational changes take time but also because the complex and tight coupling characteristics of ERP tend to make certain data quality problems unavoidable.

6. CONCLUSION

Given the essential role an ERP system plays in an organization and the importance of information and data in a knowledge-based economy, it is important that we understand its impacts on data quality.

We use NAT as a theoretical lens to examine data quality in an ERP-enabled manufacturing system. We have identified data quality problems that are difficult to overcome in this context. This is because when an ERP system is used to replace separate systems that support different functional areas, it makes these relatively loosely coupled functions within an organization more tightly coupled. Under the conditions for which the system is designed, the tight coupling effect of ERP can improve the efficiency of an organization. However, organizations often run into conditions not designed into the system. Under these abnormal conditions, the tight coupling effect reduces the effectiveness because many rippling effects caused by a mishap can quickly propagate throughout the system. Even worse, the data process in an ERP-enabled manufacturing system contains large numbers of components that dynamically interact with each other, making the system extremely complex. The systems, according to Leveson [2004], are often beyond our ability to intellectually manage: increased interactive complexity and tight coupling make it difficult for the designers to consider all the potential system states or for operators to handle all normal and abnormal situations and disturbances effectively. Thus data quality problems resulting from these two characteristics of an ERP system generally are not easy to eliminate.

Our work contributes to data quality research with a systematic investigation of data quality problems in an ERP-enabled manufacturing system. Based on our data analysis, we have developed a framework for data quality problems in such a context. This framework, presented in Figure 5, can be used to identify major data quality problems in data production, storage and maintenance, and utilization processes in ERP systems.

This study confirms the previous study on data quality. Of the 16 DQ dimensions identified in Wang and Strong [1996], our study identified eight in an ERP-enabled manufacturing system. Our study also identified DQ patterns similar to those reported in Strong et al. [1997]. For example, multiple sources of same data and judgment involved in data production are identified as the two causes of intrinsic DQ problems that involve questionable believability and objectivity [Strong et al. 1997]. This pattern also appeared in our case study. Control data such as materials data came from multiple data sources and were inspected and evaluated by a 30-person team before entering

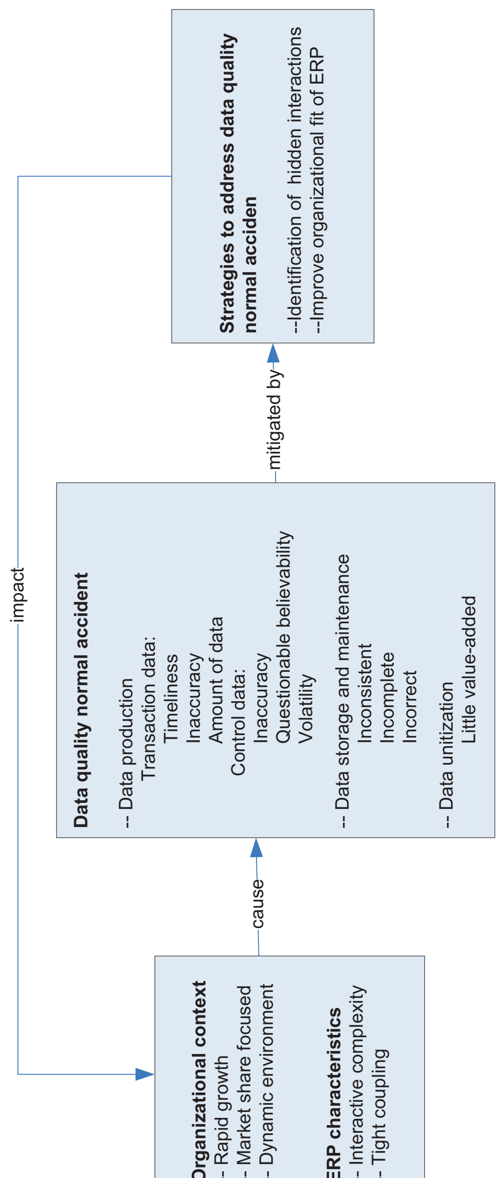

Fig. 5. Normal accidents in data quality.

into the system. Despite the effort, constant requirement changes led to inaccuracy and questionable believability in control data.

More importantly, this research extends our understanding on data quality problems in complex systems. First, while data quality has been studied in general, our study examines this issue in the context of an ERP-enabled manufacturing system. Second, prior research has identified data quality problems experienced by different user groups of ERP [Strong and Volkoff 2005]. Our work explains the *why* behind data quality problems by linking them to certain characteristics of ERP and its organizational context. Third, our study also extends prior research that identifies data quality problem patterns and suggests process-based solutions to data quality problems [Strong et al. 1997]. We have developed the normal accident perspective to understand the root causes of these problems. Fourth, we contribute to the data quality research by identifying a new data quality problem, data volatility, which has significant impacts on overall data quality in a complex and tightly coupled system.

Our research also contributes to the practice of data quality management by developing strategies to address data quality problems in ERP-enabled manufacturing systems. Previous research has shown that certain data quality problems have no easy solutions [Strong et al. 1997]. Our research, instead of focusing on solving each data quality problem separately, proposes strategies to eliminate the systemic root causes of these problems. By identifying hidden interactions to reduce the complexity of the system and improving organizational fit of ERP, our suggested strategies can address a set of data quality problems more effectively. Moreover, the use of “normal accident” theory provides us with a new perspective on data quality management within ERP systems. With this perspective, researchers and practitioners can develop data management solutions that are effective in the presence of certain “normal” data quality problems within a specific organizational context.

Our case study was conducted at one company in the manufacturing industry. The various issues revealed in our study may not represent all situations. This limitation should not prevent us from applying the findings to a broader context given the fact that an ERP is a large system similar to many other systems that tend to have normal accidents. However, applying the findings to other companies should be done with caution. For example, while ERP is a general system, the organizational context is very unique at AC Company. Whether certain data quality problems are normal accidents in a stable environment needs to be carefully evaluated.

In future research, we will examine the use of the NAT in other organizational settings. We will also test the effectiveness of the proposed strategies for addressing “normal” data quality problems. In order to improve business and IT agility, many organizations are migrating their IT infrastructure to Service-Oriented Architecture (SOA) [Choi et al. 2010; Yoon and Carter 2007]. Loose coupling is the main design principle of SOA [Mueller et al. 2010]. Our future research will also study whether the migration of ERP to SOA can effectively reduce coupling of ERP and help solve data quality problems caused by the tight coupling effect of ERP.

REFERENCES

Ballou, D., Wang, R., Pazer, H., and Tayi, G. K. 1998. Modeling information manufacturing systems to determine information product quality. *Manag. Sci. 44*, 462–484.

Ballou, D. P. and Pazer, H. L. 1995. Designing information systems to optimize the accuracy-timeliness tradeoff. *Inf. Syst. Res. 6*, 51–72.

Ballou, D. P. and Pazer, H. L. 2003. Modeling completeness versus consistency tradeoffs in information decision contexts. *IEEE Trans. Knowl. Data Engin. 15*, 240–243.

Beard, J. W. and Sumner, M. 2004. Seeking strategic advantages in the post-net era: Viewing ERP systems from the resource-based perspective. *J. Strategic Inf. Syst. 13*, 129–150.

BELOTSKY, G. AND JOHNS, H. 2008. Encapsulation and its discontents: Applying normal accident theory to software design. In *Proceedings of the Open Source Convention.*

BROOKS, F. 1987. No silver bullet - Essence and accidents of software engineering. *IEEE Comput. 20*, 10–19.

CHOI, J., NAZARETH, D. L., AND JAIN, H. K. 2010. Implementing service-oriented architecture in organizations. *J. Manag. Inf. Syst. 26*, 253–286.

CLARKE, L. 1989. *Acceptable Risk? Making Decisions in a Toxic Environment*. University of California Press, Berkeley, CA.

DAVENPORT, T. H. 1998. Putting the enterprise into the enterprise system. *Harvard Bus. Rev. 76*, 121–131.

EIN-DOR, P. AND SEGEV, E. 1978. Organizational context and the success of management information systems. *Manag. Sci. 24*, 1067–1077.

EISENHARDT, K. M. 1989. Building theories from case study research. *Acad. Manag. Rev. 14*, 532–550.

FISHER, C. W. AND KINGMA, B. R. 2001. Criticality of data quality as exemplified in two disasters. *Inf. Manag. 39*, 109–116.

GATTIKER, T. F. AND GOODHUE, D. L. 2004. Under standing the local level costs and benefits of ERP through organizational information processing theory. *Inf. Manag. 41*, 431–443.

GATTIKER, T. F. AND GOODHUE, D. L. 2005. What happens after ERP implementation: Understanding the impact of interdependence and differentiation on plant-level outcomes. *MIS Quart. 29*, 559–585.

GOODHUE, D. L. AND THOMPSON, R. L. 1995. Task-technology fit and individual performance. *MIS Quart. 19*, 213–236.

HONG, K.-K. AND KIM, Y.-G. 2002. The critical success factors for ERP implementation: An organizational fit perspective. *Inf. Manag. 40*, 25–40.

HOPKINS, A. 1999. The limits of normal accident theory. *Safety Sci. 32*, 93–102.

JACUCCI, E., HANSETH, O., AND LYYTINEN, K. 2006. Introduction: Taking complexity seriously in IS research. *Inf. Technol. People 19*, 5–11.

LALLY, L. 2002. Complexity, coupling, control and change: An IT based extension to normal accident theory. In *Proceedings of the Internationl Resources Management Conference.* 1089–1095.

LALLY, L. 2005. Information technology as a target and shield in the post 9/11 environment. *Inf. Resource. Manag. J. 18*, 14–28.

LEE, Y. W. 2003. Crafting rules: Context-reflective data quality problem solving. *J. Manag. Inf. Syst. 20*, 93–119.

LEE, Y. W. AND STRONG, D. M. 2003. Knowing-why about data processes and data quality. *J. Manag. Inf. Syst. 20*, 13–39.

LEE, Y. W., STRONG, D. M., KAHN, B. K., AND WANG, R. Y. 2002. AIMQ: A methodology for information quality assessment. *Inf. Manag. 30*, 133–146.

LEVESON, N. 2004. A new accident model for engineering safer systems. *Safety Sci. 42*, 237–270.

MACKENZIE, D. 1994. Computer-related accidental death: An empirical exploration. *Sci. Public Policy 21*, 233–248.

MARKUS, M. L. 1983. Power, politics and MIS implementation. *Comm. ACM 26*, 430–444.

MASON, J. 2002. *Qualitative Researching.* Sage Publications.

MILES, M. B. AND HUBERMAN, A. M. 1994. *Qualitative Data Analysis*. Sage Publications.

MINTZBERG, H. 1979. *The Structuring of Organizations*. Prentice Hall, Englewood Cliffs, NJ.

MUELLER, B., VIERING, G., LEGNER, C., AND RIEMPP, G. 2010. Understanding the economic potential of service-oriented architecture. *J. Manag. Inf. Syst. 26*, 145–180.

NEUMANN, P. G. 1995. *Computer Related Risks*. ACM Press, Addison Wesley Publishing Co., New York.

ORLIKOWSKI, W. J. 1993. CASE tools as organizational change: Investigating incremental and radical changes in systems development. *MIS Quart. 17*, 309–340.

PARÉ, G. 2004. Investigating information systems with positivist case study research. *Comm. AIS 2004*, 233–264.

PATTON, M. P. 1990. *Qualitative Evaluation and Research Methods.* Sage, London.

PAWLOWSKI, S. AND ROBEY, D. 2004. Bridging user organizations: Knowledge brokering and the work of information technology professionals. *MIS Quart. 28*, 645–645.

PERROW, C. 1999. *Normal Accidents: Living with High-Risk Technologies*. Princeton University Press, Princeton, NJ.

RAYMOND, L. 1990. Organizaional context and information systems success: A contingency approach. *J. Manag. Inf. Syst. 6*, 5–20.

SAGAN, S. 1993. *The Limits of Safety: Organizations, Accidents, and Nuclear Weapons*. Princeton University Press, Princeton, NJ.

SAMMARCO, J. J. 2003. Addressing the safety of programmable electronic mining systems: Lessons learned. In *Proceedings of the IEEE Industry Applications Conference, 37th IAS Annual Meeting.*

SCHULTZ, R. L. AND SLEVIN, D. P. 1975. Implementation and organizational validity: An empirical investigation. In *Implementing Operations Research / Management Science,* R. L. Schultz and D. P. Slevin, Eds., Elsevier, 153–182.

SHORT, J. F. 1984. The social fabric at risk: Toward the social transformation of risk analysis. *Amer. Sociol. Rev. 49*, 711–725.

SIRKISOON, F. AND SHEPHERD, J. 2002. The steady stream of ERP investments: AMR research. http://www.amrresearch.com/Content/View.asp?pmillid=14775.

SOH, C., KIEN, S. S., AND TAY-YAP, J. 2000. Culture fits and misfits: Is ERP an universal solution? *Comm. ACM 43*, 47–51.

SOMERS, T. M. AND NELSON, K. G. 2004. A taxonomy of players and activities across the ERP project life cycle. *Inf. Manag. 41*, 257–278.

STRAUSS, A. AND CORBIN, J. 1990. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Sage Publications, London.

STRONG, D. M., LEE, Y. W., YANG, W., AND WANG, R. Y. 1997. Data quality in context. *Comm. ACM 40*, 103–110.

STRONG, D. M. AND MILLER, S. M. 1995. Exceptions and exception handling in computerized information processes. *ACM Trans. Inf. Syst. 13*, 206–233.

STRONG, D. M. AND VOLKOFF, O. 2005. Data quality issues in integrated enterprise systems. In *Proceedings of the 10th International Conference on Information Quality*. F. Naumannn, M. Gertz and S. Madnick, Eds., 126–132.

THOMPSON, J. D. 1967. *Organizations in Action*. McGraw-Hill, New York.

VAUGHAN, D. 1996. *The Challenger Launch Decision. Risky Technology, Culture, and Deviance at NASA.* The University of Chicago Press, Chicago, IL.

VENKATRAMAN, N. 1994. IT-enabled business transformation: From automation to business scope redefination. *Sloan Manag. Rev. 35*, 73–87.

VOLKOFF, O., STRONG, D. M., AND ELMES, M. B. 2005. Understanding integration effects of enterprise systems. *Euro. J. Inf. Syst. 14*, 110–120.

VOSBURG, J. AND KUMAR, A. 2001. Managing dirty data in organizations using ERP: Lessons from a case study. *Industr. Manag. Data Syst. 101*, 21–31.

WANG, R. Y. 1998. A product perspective on total data quality management. *Comm. ACM 41*, 58–63.

WANG, R. Y., LEE, Y. W., PIPINO, L. L., AND STRONG, D. M. 1998. Manage your information as a product. *Sloan Manag. Rev. 39*, 95–105.

WANG, R. Y. AND STRONG, D. M. 1996. Beyond accuracy: What data quality means to data consumers. *J. Manag. Inf. Syst. 12*, 5–34.

WEICK, K. 1990. Technology as equivoque: Sense-making in new technologies. In *Technology and Organizations*, P. S. Goodman and L. S. Sproull, Eds., Jossey-Bass, Stanford, CA, 1–43.

WEICK, K. 1995. *Sensemaking in Organizations*. Sage Publications.

WEICK, K. 2001. Enacted sensemaking in crisis situations. In *Making Sense of the Organization,* K. Weick, Ed., Blackwell, Cambridge, MA, 224–236.

WEICK, K. 2004. Normal accident theory as frame, link, and provocation. *Organ. Environ. 17*, 27–31.

WEIL, P. AND OLSON, M. H. 1989. An assessment of the contingency theory of management information systems. *J. Manag. Inf. Syst. 6*, 59–85.

XU, H., NORD, J. H., BROWN, N., AND NORD, G. D. 2002. Data quality issues in implementing an ERP. *Industr. Manag. Data Syst. 102*, 47–58.

YIN, R. K. 2003. *Case Study Research: Design and Methods*. Sage Publications, London.

YOON, T. AND CARTER, P. 2007. Investigating the antecedents and benefits of soa implementation: A multi-case study approach. In *Proceedings of the Americas Conference on Information Systems (AMCIS'07).*

Received February 2010; revised August 2010; accepted February 2012