Managing dirty data in organizations using ERP: lessons from a case study

Authors: Jodi Vosburg
Authors: Anil Kumar

Volume: 101
Issue: 1
Pagination: 21-31
ISSN: 02635577

Subject Terms: Data collection
Data integrity
Enterprise resource planning
Systems management
Studies
Information technology

Abstract:

The integrity of the data used to operate and make decisions about a business affects the relative efficiency of operations and quality of decisions made. Protecting that integrity can be difficult and becomes more difficult as the size and complexity of the business and its systems increase.

Recovering data integrity may be impossible once it is compromised. Stewards of transactional and planning systems must therefore employ a combination of procedures including systematic safeguards and user-training programs to counteract and prevent dirty data in those systems. This paper discusses issues related to the origin of dirty data, associated problems and costs of using dirty data in an organization, the process of dealing with dirty data in a migration to a new system: enterprise resource planning (ERP), and the benefits of an ERP in managing dirty data. These issues are explored in the paper using a case study.

Copyright MCB UP Limited (MCB) 2001

Full Text:

Jodi Vosburg: The University of Wisconsin-Whitewater, Wisconsin, USA
Anil Kumar: The University of Wisconsin-Whitewater, Wisconsin, USA
Daily operations, planning, and decision-making functions in organizations are increasingly dependent on transaction data. This data is entered electronically and manually and then organized, managed and extracted for decision-making.

The same data entered and used to facilitate building, shipping, and invoicing goods is also extracted and manipulated to evaluate factory and sales force performance in the short term. In the long term this data is used to chart the course of the business in terms of manufacturing facilities, products, and marketing. The integrity of the data used to operate and make decisions about a business affects the relative efficiency of operations and quality of decisions made. Protecting data integrity is a challenging task. Redman (1995) comments that "many managers are unaware of the quality of data they use and perhaps assume that IT ensures that data are perfect. Although poor quality appears to be the norm, rather than the exception, they have largely ignored the issue of quality". Other scholars (Greengard, 1998; Kilbane, 1999; Tayi and Ballou, 1998; Wallace, 1999) also point out the importance of data quality for organizations.

Maintaining the quality of the data that is used in an organization is becoming an increasingly high priority for businesses. In a recent survey of 300 IT executives conducted by Information Week (Wallace, 1999), majority of the respondents (81 per cent) said, "improving customer data quality was the most important post-year 2000 technology priority". The respondents further stated that there would be "significantly increased spending" on data quality in their organizations. Companies that manage their data effectively are able to achieve a competitive advantage in the marketplace (Sellar, 1999). On the other hand, "bad data can put a company at a competitive disadvantage" comments Greengard (1998). A recent study (Ferriss, 1998) found out that "Canadian automotive insurers are taking a major hit from organized and computer-literate criminals who are staging crashes and taking advantage of dirty data in corporate databases". The study found out that in one case several insurance firms lost $56 million to one fraud ring.

How does a company end up with dirty data and what can be done to prevent this? Disparate data stores (individual, departmental, and organizational) that have been developed and used by organizational users over the years lead to dirty data problems. For example, dissimilar data structures for the same customer data (spelling discrepancies, multiple account numbers, address variations), incomplete or missing data, lack of legacy data standards, actual data values being different from meta-labels, use of free-form fields, etc. (Kay, 1997; Knowles, 1997; Weston, 1998). These problems can be compounded by the volume of data that is stored and used in organizations. One way of overcoming this problem is to use technologies that integrate the disparate data stores for an organization and help companies clean up their data. Enterprise resource planning (ERP) systems (SAP, Peoplesoft, Baan, J.D. Edwards, etc.) are examples of such systems. "A good ERP system offers an integrated option, implementing browser and client-server modes while maintaining consistent data and function within the enterprise and out to the supply chain" (Stankovic, 1998). In recent years, ERP vendors have gone beyond providing the traditional integrated applications, such as manufacturing, financials, and human resources. Newer applications that have emerged include supply chain management, customer-relationship management, data mining and data warehousing (Caldwell and Stein, 1998; Stankovic, 1998) and browser modes that enable organizations to reach out to customers and the supply chain. Caldwell and Stein (1998) also point out
that "most important, ERP forces discipline and organization around processes, making the alignment of IT and business goals more likely in the post-ERP era". Aligning IT and business goals has always been a top priority for senior management. Thus it might be helpful for a company to implement an ERP system.

In this paper, we discuss the experiences of a company, which implemented an ERP system in their organization. The discussion is focused primarily on the data aspect of the implementation. The paper is organized as follows. In the next section we describe the case-study organization. Section 3 defines the concept of dirty data and its impact on the integrity of organizational data. In Section 4 we list the costs incurred by organizations as a result of using dirty data. Section five highlights several lessons learnt from the case-study organization and, finally, in Section 6 we summarize the guidelines for companies planning to implement ERP solutions to overcome dirty data problems.

2.0 The case study

The organization where this case study was conducted is a $650 million division of a Fortune 500 company located in the Midwest. This company is a manufacturer of electrical, lighting, and automotive equipment. The products of this company are marketed domestically and internationally. The company employs approximately 1,600 people in manufacturing and sales facilities located both domestically and internationally. There are 17 manufacturing facilities located in North America and Asia. The case study was used to understand the implications of dirty data at the company before and after the implementation of an ERP system. The ERP implementation in the company replaced a number of independent mainframe legacy systems used for order and quotation processing, manufacturing, transportation, billing, and finance applications. One of the co-authors of the study works at the company as the system/support supervisor for the Customer Support Center (CSC). In this role, the author was directly involved in identifying, trouble-shooting, and training for dirty data concerns in data entry and with specifying, testing, and distributing customer and sales-force reports. In addition, we interviewed several other employees in the organization who were involved with this project. These employees included the manager of the CSC and marketing services, an information analyst in the marketing services group, and a customer support representative (CSR). The manager of the CSC is responsible for managing domestic order processing and sales and marketing reporting for the division. The information analyst works with users and programmers to specify report requirements and does much of the testing and trouble-shooting for those reports. The CSR is the data entry point analyzing and translating customer purchase orders into ERP documents. This study will look primarily at issues relating to the CSC.

3.0 Dirty data defined

"At first, the abbreviation for black was blk. Then it was changed to bck. We didn't discover this change until someone said the color mix didn't look right (Horwitz, 1998)."

Dirty data exists when there are inaccuracies or inconsistencies within a collection of data or when data extraction is inconsistent with intent. Inclusion of dirty data in a data source may pollute the entire data source making it difficult or unwise to use the data for analysis.
Dirty data in a transactional system can mean incorrect order taking, products not built to specification, or errors in packaging, documentation, or billing.

The result is dissatisfied customers, loss of shareholder confidence, unnecessary material and labor costs, and the real and opportunity costs of time spent correcting errors resulting from dirty data. Those interviewed define dirty data as follows: "The GIGO (garbage in, garbage out) theory applies to dirty data. If you don't have checks in the system that prevents human error, you will have errors in your data. Data integrity refers to data that is systematically edited or edited by "experts" after data entry to remove errors (Manager, CSC)." "Duplicate data or data that is incomplete or extraneous (Information Analyst, Marketing Services)." "Anything that is entered incorrectly (CSR)."

The definitions used reflect each one's experience with dirty data. Awareness of this problem is growing within the organization as users, systems people, and management uncovers and deals with problems resulting from dirty data.

Data integrity requires awareness and control of dirty data. A collection of data has integrity if the data are logically consistent and accurate. Data integrity requires that data additions or changes be reflected in each of the locations where that data is stored and that data is consistent across the storage medium(s) used. Data integrity also requires that the users of that data understand the meaning of the data within the context of the business. Maintaining data integrity requires a systematic approach to data processing, storage, sharing, manipulation, and reporting.

4.0 Cost of using dirty data

"Errors in data can cost a company millions of dollars, alienate customers, and make implementing new strategies difficult or impossible" (Redman, 1995). The manager of CSC commented that:"Any business that has to issue debits and credits or that throws out surplus, unusable inventory, understands the costs of dirty data. Each credit or debit is estimated to cost the company $75 for the clerical efforts of analyzing, generating and disseminating the document. Added to that are the following: production errors from erroneous bills of material or misinterpretation of a customer’s specifications; freight costs for shipping and returning product; inventory scrapping charges where the product cannot be resold; financial penalties charged by the customer for our error; ordering of unneeded materials; scrapping of raw materials; wasted labor charges at the organization and its customer; warranty charges to fix the product, if it can be modified; and unknown cost of the customer not ordering additional product from you because of your data problems. The managers and people involved in warranty, credit and collection and finance understand the ramifications. The rest of the organization understands what their managers or supervisors have shared with them. Our quality program emphasizes feedback to the person involved with a quality problem. It is up to the management team to insure that all people understand the problems dirty data can cause as well as prevention."

The information analyst for marketing services was of the opinion that "most of the costs associated with dirty data cannot be measured in terms of dollars. If these costs could be quantified the management would be shocked". She stressed the cost of the endless number of consultants required to configure the system to prevent a particular data problem or to determine or correct the results of one.
The CSR focused on "costs associated with customer dissatisfaction - lost confidence and business are hard to measure and harder to win back". She pointed out the frustration and time lost at the factory, in the marketing departments, and at the CSC in correcting problems resulting from dirty data.

Each person's perspective is culled from that person's training and experience. The CSR indicated that she had little understanding of the way in which the data she enters is used in peripheral departments and how it becomes part of reporting. For that reason, it is important to examine the data and rationalize it. Data rationalization involves determining what data is important to which department and prioritizing the value of those data sets. Once this determination is made, plans to correct and prevent dirty data can be laid.

5.0 The ERP implementation: lessons learned

The start of data integrity problems is really a failure to treat data as a strategic business resource. Scholars (Redman, 1995; Tayi and Ballou, 1998) point out that data is a key organizational resource. However, as pointed out by Kilbane (1999), "Many companies who use data contained in legacy systems are not leveraging it as a strategic company asset." The primary challenge to maintaining data integrity is the lack of resources allocated to it. To maintain data integrity, people with an understanding of the origins and results of dirty data and the ways to prevent and correct it, must be dedicated to the task. Redman (1995) says that: "Due largely to the organizational politics, conflicts, and passions that surround data, only a corporation's senior executives can address many data quality issues. Only senior management can recognize data (and the processes that produce data) as a basic corporate asset and implement strategies to proactively improve them." Where data integrity is one of many responsibilities of people with no understanding of the concepts surrounding data integrity, dirty data is the result. Integrity, issues receive attention in times of crisis, but as soon as the crisis is over, those with responsibilities other than data integrity turn to the pressing deadlines or daily tasks that they are responsible for. In a complex ERP environment, this can result in perpetual crisis management. In the following paragraphs we discuss the lessons learned from this case study.

5.1 Understanding and communicating new demands of an ERP system

Before the move from legacy applications to an ERP system takes place, considerable thought should be given to how the system change will change the roles of the users. The conversion to an ERP system is not just a data extraction, cleansing, transformation, and populate process to effectively implement an ERP system. An organization needs a strategy and a plan. Atre (1998) points out that "legacy data is invariably in worse condition than you realize". Caldwell and Stein (1998) comment that "ultimately, by feeling their way through the initial shock of an ERP implementation - new business processes, new job roles, new management structures, and new technologies - companies are transforming themselves".
In this company there are 48 CSRs in the CSC. These CSRs are responsible for entering orders taken from domestic customers. Now, with the ERP, the items on these orders not only initiate the manufacturing, shipping, and invoicing functions, but also are the raw data used to generate the sales and marketing reports. The sales and marketing reports feed the decision-making processes that steer the business. The correct and consistent entering of these orders is critical to preventing dirty data.

Most CSRs believe that the order entry process has increased in complexity with the implementation of the ERP. Some estimate that the time required to enter an order has increased two to four times. The reasons for this widely-held perception are threefold. First, the ERP is still quite new - system glitches can mean several unsuccessful attempts at entering a single order and the eventual involvement of system support personnel in processing. Second, there are more steps to the order entry processes, and greater variation across product lines. Legacy systems were used for narrowly-defined transaction sets. For example, each of the four product groups in the company had their own manufacturing system. The homogeneity of the transactions and of the users meant that the legacy systems could be customized to accommodate those tasks without affecting the ability of other users to perform other tasks. Now that all users share a single system, transactions must be generalized to fit all tasks. Where customization cannot be automated, it becomes a manual part of user work processes - the order entry process varies greatly from product line to product line. Greater expertise is required on the part of the user, not only in the performance of their assigned tasks but also in those of others that are affected by their system transactions. The learning curve has been steeper than anyone imagined. Third, data entry skills are no longer enough to successfully enter orders- the ERP requires system savvy and an analytical approach. It has become critical that CSRs understand the logic behind the processes and the ramifications of their actions on-line. Many are inexperienced in this way of working. The combination of these factors has increased the occurrence of inaccurate, inconsistent data being entered on the ERP via sales orders, as CSRs attempt to complete their complex and time-consuming data entry work in the same amount of time they did prior to the ERP implementation and without a clear understanding of how that data is to be used by other functional areas in the business and by upper management for analysis and business decisions. "Lesson: Organizational users need to be educated and prepared for the changes that will take place as a result of ERP implementation."

5.2 Developing shared understanding of data

The lack of a shared understanding of the uses and value of data among those performing the same tasks and among those performing different tasks can lead to creation of dirty data. Tayi and Ballou (1998) point out that "the data gatherer and initial user may be fully aware of the nuances regarding the meaning of the various data items, but that will not be true for all of the other users". Where those performing the same tasks have a different understanding of the data being processed, inconsistencies are inevitable. For example, if the marketing services department members differ on whether abbreviations are to be used in customer master data, inconsistent entry is the result. Locating this data becomes difficult for CSRs because they cannot be sure if they are using the wrong abbreviation, or if the data has not been entered. The result of this lack of shared understanding is duplicate records - when the CSR cannot find the record that they are looking for, a new record is requested. Even if marketing services is able to locate the record and corrects
the abbreviation before creating a duplicate record, both the CSR and marketing services have spent unnecessary time.

A lack of a shared understanding is common among data generators and report writers. A CSR knows that the promised ship date on an order with a production block is not valid, but a consultant writing a backlog report probably does not. As a result, the invalid date is published on the report. Geographical distances and functional barriers exacerbate this complexity. The further an employee is from another employee, and the less that employee understands what is required in the other's position, the less likely they are to share a common understanding of the importance of the data each deals with. According to the CSC manager: "In the business right now, those entering the data and those using the data are so confused that there is little understanding of the data in the system. We are working with users AND the IT departments to share the knowledge about the entered, calculated AND extracted data. Without this, we are, and have been, subject to interpretation of a field with a title meaning different things to those entering versus using the data. We are finding how difficult it is to deal with a program written in another language, as field translations have always assisted users and IT people in the past. In our ERP, there is no such extra help available for those looking for field definitions and understandings." "Lesson: Champions of the ERP implementation project should ensure that all users understand the organizational data in a manner that is consistent throughout the organization."

5.3 Ownership of data and responsibilities

Responsibility for ensuring data integrity belongs to all employees. Tayi and Ballou (1998) comment: "The capability of judging the reasonableness of the data is lost when users have no responsibility for the data's integrity and when they are removed from the gatherers." Atre (1998) points out: "IT staff need help and cooperation from business users to identify and cleanse operational data. Users should be primarily responsible for determining the business value of data. Don't rely on systems integrators - they don't understand the business value of the data." One has also to consider the "politics" which play an important role. Often managers may agree that they own the data, but may want everybody to be involved in cleaning it. The manager of the CSC commented:"I believe that data integrity is the responsibility of every company employee. All positions, all departments are responsible for insuring the data they are entering, reviewing or utilizing is error free. It is the responsibility of every manager to make sure the tools are in place to insure data integrity for the data they are responsible for.... In the past, users relied on the IT departments to make sure the edits were in place to make the data correct. With ERP systems and more user controlled systems and input, it is a joint responsibility. Users must understand systems better and IT personnel must understand business problems better in order for them to work together to achieve the highest level of data integrity. Too many IT people are good programmers but not good business analysts."

The marketing services staff is responsible for maintaining customer and salesforce master data, for testing and maintaining the ERP data structures that define transactional data, and for authoring and generating sales and marketing reports. Marketing services has been successful in guarding against duplicated records, misspelled names, inverted text, missing fields, outdated area codes and ZIP codes, and other kinds of dirty data in customer and salesforce master data by employing combination of user training, well
defined procedures, and tight control and auditing of additions, changes, and deletes. Every CSR received at least four hours of training on the use and import of this master data. During that training, CSRs were asked to review the master data for their assigned customers and to advise marketing services of any necessary changes. New master data requires the completion of a form to ensure all necessary information is provided. Only two people in the marketing services department do the actual addition of the new data to ERP. In addition, an audit report is run regularly to identify changes made to the data. This report helps to catch mistakes and identify where additional training is required. A data steward, who is responsible expressly for protecting data integrity, should support the efforts of the CSRs and the marketing services department. This data steward would be responsible for raising awareness about data issues and implementing systematic procedures for data auditing and user training. "Lesson: Ensuring that all stakeholders of an ERP system understand their responsibilities with respect to maintaining data integrity will lead to a better quality system. Data that is a part of an ERP system belongs to an organization and not to any individual department or user."

5.4 Migrating legacy data

Ruber (1999) comments, "Migrating information from departmental databases and transaction-processing systems ... is a daunting task." He goes on to say that the "hardest part is cleansing the data, yet people tend to underestimate that part of the process." Legacy systems in corporations, which were created in different generations, create major stumbling blocks for migrating data to integrated application systems. Quick fixes that become embedded in the case of legacy systems create complexities that are difficult to overcome. Most of these systems are usually lenient with the data that is maintained, resulting in lack of data standards or documentation in the form of metadata. Before this data is migrated there is a need to clean it. An effective strategy for companies planning to implement integrated applications, such as ERP, may be to use automated tools for cleaning legacy data before integrating it. Tools provided by vendors such as id.Centric, Vality Technology, HarteHanks, etc. (Knowles, 1997) may benefit organizations significantly. Sales of such tools used for data extraction, refining and loading, was expected to reach $210 million by the end of 1999 (Kay, 1997).

The initial ERP implementation involved a programmatic load of legacy sales order backlog onto the ERP. The order load program was developed and tested over a period of months by a programmer familiar with the organization's business practices and a team of users. The load was simplified by the fact that the legacy system was well supported. That support meant not only that data to be converted was relatively clean, but also that the data in the legacy system was well-defined and understood - a program could be written to capture only relevant data. Unfortunately, the idiosyncrasies of the order entry process for the various product lines and of the ways in which CSRs entered the orders meant that no program could convert the data without some errors. Because the data integrity of the orders was so important, each converted order was reviewed on the ERP by the responsible CSR. Many data errors were caught and corrected during this review, including item quantity, material number, and ship to errors. But some were missed. A tremendous amount of time has been spent and is still being spent to correct these errors. One of the most common errors involved contract release orders. The program designed to convert the data somehow selected and input the wrong material number into the converted order. The CSRs, possibly tired after consecutive 12-hour days of data verification, missed many material errors. These kinds of errors, though, are always found
eventually - usually by the customer. The results of these errors were: shipment of the wrong materials; angry customers; time spent investigating the error; cost of processing credit orders and replacement orders; expedited production of the correct materials (resulting in late shipments of other orders); transportation costs for returning the wrong units; and/or cost of scrap or storage. No attempt has been made to assign a dollar value, though, to the results of this dirty data.

Overall, though, this data migration was a success - the inevitable data errors were identified, some sooner, some later, and corrected. The success was due in large part to the fact that the legacy system was well supported, the migration process was well tested and documented, and those closest to the data verified the data after the migration.

Customer master data was loaded programmatically initially. Customer master data includes addressing for billing and shipping, tax identification numbers and designations of customer type and pricing levels. Migration from legacy systems to the ERP has allowed marketing services an opportunity to scrutinize and clean data maintained about our customers and salesforce. More stringent master data requirements in the ERP, in fact, made this a necessity. For example, the legacy system had used an address in Varnons, Georgia, for one customer for years. This address was kicked out in the programmatic load of customer master data. On investigation, it was discovered that Varnons was not a city, but a stop on a railway line. The ERP will determine the zip code and county given a city and state. This not only ensures that the city and state are entered accurately, but also that the customer has provided a valid city and state combination. Customer data moved from the legacy to the ERP system were relatively cleaner than they had been on the well-maintained legacy system.

Migrating poorly-maintained legacy data

Atre (1998) comments that "you are likely to run into problems such as incompatible data formats, codes that no one can decipher, data that's embedded in long text fields, overlapping customer records from multiple systems, some with redundant data and others with conflicting or outdated data and even chunks of mystery data of long-forgotten provenance and uncertain ownership". Weston (1998) suggests using flags for dirty data that is migrated. As a result, a decision-maker can decide if he/she wants to use the information or leave it out during data analysis activities. The customer and salesforce master load for the migration from a much less well-maintained system was a more tedious and difficult procedure. Keeping that data clean on this legacy system was never a priority. The data was entered by the order entry group, as there was no position assigned to the management and control of master data for this system. Misspellings, duplicate records, and inconsistencies were the result of a lack of control over who could add, change, or delete customer master data, of instructions for proper management of the data, and of auditing procedures. The problems were exacerbated by the fact that, when the company purchased this facility, a completely new group of users began to enter this data. A lack of shared definitions of the components of the master data and their uses increased the number of discrepancies and errors. Where the original group might define a salesperson as a customer or a vendor or an agent, the company group defined the salesperson as an agent only. Subsequently, where there was no agent-type record for a particular salesperson, one was created, thereby creating the potential for inaccurate reporting of sales data. Thus, before any master data could be moved to the ERP, each
record had to be manually reviewed. The marketing services group again handled this process. Spelling errors, duplicate records, and incomplete data were addressed before the data was loaded to the ERP.

The sales order and production data on this system had been subject to inexplicable changes. For example, in November of 1998, the order entry group started to notice that some items on orders were being closed by the system for no apparent reason. Thus, they would never be built or shipped. The in-house support could not identify the cause or propose a solution, nor could the manufacturer of the software. The in-house support group advised CSRs to address these system-generated cancellations as they happened - a virtually impossible task. After much discussion, the support team agreed to write a report to locate these items.

The data on this legacy system was not well supported or understood. The data was in such poor condition that sales and marketing reports generated from system data were virtually useless. For example, the Canadian order entry location might enter orders using the same customer master record for different customer locations by overwriting the sold-to-address text to reflect the different location addresses. The domestic location would add new customer master records for each customer location. Existing reports could not accurately reflect these contradictory approaches.

These factors combine to make a programmatic migration of the sales order data to the ERP infeasible. Instead, sales orders were manually loaded onto the ERP by CSRs using an expanded backlog report. The lack of understanding of the way the system stores data, coupled with inaccuracies and inconsistencies in order entry and processing, made the writing of this report very difficult. For example, the initial run of the report included thousands of freight items. Freight items are added to sales orders by the shipping department to indicate carrier and shipment date of materials on the order - they are not backlogged. These were difficult to suppress in the report because of the inconsistent ways in which they have been added to the sales orders - some were loaded as text items, some as freight items, and some in a closed status, some in an open status. Attempts to suppress this data on the conversion order might have, without extensive testing, resulted in inadvertent suppression of materials that should be converted - the Miami order entry location uses freight items not to communicate shipment information, but to charge the customer.

This project, though, was also a success. While the manual conversion presented an opportunity for entry error, the process was largely error free. This can be attributed to the extensive testing of the backlog report serving as the basis for the conversion, simple comprehensive check-list style instructions for the CSRs in the use of the backlog report, and, most importantly, a group of CSRs now more comfortable and experienced in the use of the current ERP. Again, migration to the new ERP was a boon, because it drove the process of examining and cleaning current data. "Lesson: Migrating dirty data is a challenging task. Use of automated tools is a good strategy for organizations planning to implement integrated application systems. The most important factor is that the data needs to be cleaned before it is migrated to an ERP system."

5.5 Recognizing the complexity of integrated data
The integration of several business functions on a single system holds tremendous potential for reporting. All transactional data is now available from one source. Reporting that was difficult or not feasible in the past is now possible. This consolidation of functions onto one system has forced the various units of the business to develop a greater understanding of the work done by other units of the business and their interpretation of the data. With this potential, though, comes increased complexity. Tayi and Ballou (1998) point out "personnel databases situated in different divisions of a company may be correct but unfit for use if the desire is to combine the two and they have incompatible formats". Kilbane (1999) says "the problem is that data is, too often, in different formats and companies don't know how to properly bring it together and turn it into actionable information".

Locating data tables within the ERP system appropriate for the intended reporting has turned out to be more tedious and difficult than anyone imagined. Reports used by the salesforce and in manufacturing to describe sales order backlog have been found to be so error-ridden that they have been totally scrapped and rebuilt. Reports meant to describe incoming businesses took months to write. Several iterations of these reports were developed before the set currently in circulation was completed.

The information analyst describes an error that she stumbled across while researching another reporting data discrepancy. It seems that the same incoming business report was run for the month of March on April 1 and then again on April 3. She noticed that the totals were different. This should never occur - once the month is closed, no updating should occur. She indicates that locating the cause of a problem like this is difficult and time consuming and sometimes proves to be impossible. In this case, though, they were able to locate the source of the problem - the reporting structure was identifying the wrong date field as the determinant for which month a particular type of order would be allocated to. The correction of this structure error is perhaps more tedious than finding the cause of it - the field reference must be changed in more than 100 places in each of the several data structures. These and other integrity problems detected in the early going have meant that several manual adjustment schedules must be published with each run of this report - the data cannot be cleaned. The information analyst attributes these errors to a lack of comprehensive testing of the updating that occurs when these orders are processed. She sights a lack of communication between those that understand the way the company accrues and processes data and those responsible for building the data definition structures. As a result, some basic assumptions were made in the definition of data that were incorrect.

The complexity entailed by system integration is compounded by the marketing services staff's inexperience with the selected reporting bolt-on, the ERP data structures, and the architecture of the data itself. Basic reporting requirements to operate the business, coupled with this inexperience, have resulted in an inordinate reliance on consultants for report writing. While these consultants are skilled in report writing and the integration of ERP, their lack of understanding of company business and the transactional data and processes, and subsequent ERP configuration changes, has impeded accurate reporting."Lesson: It takes time for users to comprehend and use integrated data as a result of using ERP packages. Care should be taken to ensure that all users understand the concept of integrated corporate data and use it accordingly."
5.6 Testing the new system

The costs of insufficient testing prior to implementation can be very high. Months of suspicions were confirmed in March, when it was discovered why incoming business numbers seemed too high. CSRs had been entering the sales credit designation on orders more than once. Whenever a new item is added to an existing ERP sales order, the ERP returns an error indicating that sales credit is missing. The correct action is to activate the existing sales credit designation on the order for the new items. This problem was not anticipated or clearly understood. Thus the correct handling was never made part of the ERP training for CSRs. So, CSRs generally entered an additional sales credit designation with each addition to an order. Some orders showed a sales credit allocation of 400 per cent or more of the net value of the order. The sales credit numbers are also used to report incoming business. In total, this data entry error resulted in an eight million-dollar overstatement of incoming business. Because this affected incoming business and not shipments or production, the cost was minimal financially. However, sales managers were forced to adjust sales engineer bonuses downward as a result of the discovery.

This data cannot be corrected on the ERP system. All adjustments had to be handled manually. Some preventative measures were immediately put in place. In the short term, additional training was provided to the people who enter orders to make them aware of the impact of this error. A daily report is being run to identify these errors as they are made, allowing on-line corrections. In the long term, the ERP configuration changes have been requested to eliminate the misleading error message and to add messages when more than 100 per cent of the value of the order is allocated as sales credit. According to the manager of the CSC: "The problem might have been prevented if we all knew how to test wrong. In all the massive testing done on order entry and reporting on it, not enough was done to try to enter bad data. Some of the edits seemed so self-evident, that there lack was almost impossible to comprehend. I think we are just now learning how important understanding and testing for dirty data is in a truly integrated system.""Lesson: Test, test and test again. Testing is a crucial aspect of implementing ERP solutions. There should be no short-cuts in testing. Different user groups should be involved in the testing process to ensure that all possible scenarios are used for testing the ERP system before the conversion to ERP is implemented."

5.7 Training

Lack of proper training can frustrate users when they begin using an ERP system in an organization. Caldwell and Stein (1998) point out the example of Amoco, where "managers found SAP so unfriendly they refused to use it. Few [of our] people use SAP directly because you have to be an expert". The authors further comment that in the case of Owens Corning, the organization found out that "the cultural and organizational impact on IT organizations is a little short of revolutionary". The entry and extraction of dirty data can be prevented with greater dedication to initial and on-going training for those responsible for entering and extracting data. A lack of time is typically sighted as the reason for inadequate training. The time required to investigate, understand, correct, and prevent problems due to dirty data is considerably more, though, than that required simply to
understand and prevent those problems. The additional cost of this reactive approach is the loss of shareholder confidence in the system, employees, and data.

A significant training effort was put into teaching those that would be using and entering data in the system. Each CSR received in excess of 50 hours of training in meaning and population of the various fields comprising the order entry screens. Order entry procedures are documented in detail and available to all CSRs. The difficulty lies in knowing how much training is enough - a difficult question to answer at conversion time, given the consultants' lack of understanding of the particular business and the employees' lack of understanding of the new system and the potential problem areas. There is no question, though, that additional training will be required after implementation to address the numerous unanticipated problems that will arise. "Lesson: On-going training is a prerequisite for success in implementing ERP systems. Organizations should plan ahead of time to train all users before and after the implementation. Periodic exchange of ERP experiences by users in an organization from their work environment will go a long way."

5.8 Prioritizing data maintenance

According to the CSC manager: "Data integrity is assigned a high priority at the management and IT level. It is not as high a priority at the middle manager and lower levels, as worrying about data integrity can slow down production, order entry, shipping, etc."

The information analyst indicated that data integrity was critical in the marketing services department, but prioritized much lower in departments dealing with the day-to-day operations.

Even marketing services, though, does not have a system in place to check data regularly for problems. The information analyst indicates that the department spends so much time "putting out fires" that there is little time left over for carrying out systematic data checks. The problem is exacerbated by a lack of tools for auditing. The information analyst attributes this to the newness of the implementation.

At present, data integrity is protected through a combination of system safeguards, user training, and data entry procedures. System safeguards are the result of building data integrity rules into ERP. For example, ERP will prevent a CSR from entering a ship to address in a sold-to-field. This is a hard error, preventing saving of the data. Soft error messages give the CSR the opportunity to review potentially erroneous data. Additionally, many fields are populated from drop-down boxes, eliminating the chance for misspelled entries or entries outside the acceptable domain for the field. "Lesson: Organizations should emphasize that maintaining data integrity is an on-going process and everybody needs to play an active role. Maintaining data integrity does not stop with the implementation of the ERP system."

5.9 Using consultants

Care must be taken to ensure that if consultants are hired for the transition project, the internal stewards of the system understand their work. For example, in this company, consultants were responsible in large part for developing data structures for the new
system, and form the system metadata. These structures are used in conjunction with raw data to define the context of the data and to ensure that data reported is consistent with what is intended or required. For example, a structure may define incoming business as the value of the selling prices on sales orders not including items that have been cancelled. Thus reports providing incoming business data will not include cancelled items. Direct involvement by the manager of the CSC and the marketing services staff throughout the development, ensured that data structures defined by the consultants matched data the way users of that data defined it. This prevents the possibility that once the consultants leave the project, the users of the system understand the data that is being processed by the system. “Lesson: Hiring consultants to assist with the ERP implementation is an effective strategy if organizations ensure that all work done by the consultants is understood and documented. The ERP implementation knowledge should not leave the organization after the consultants work is completed.”

5.10 Post-ERP implementation

Counteracting and preventing dirty data - current perceptions and practices

Data entry procedures have been created to control the potential damage accruing to dirty data. For example, CSRs are required to place a production block on orders for some material types. This gives the product group marketing departments an opportunity to review the order and correct any errors before production begins. Taken individually, this procedure seems like a reasonable safeguard. Taken together with all of the other exceptions and qualifiers to the basic order entry procedure based on material type, or product line, and order type, though, the procedures begin to seem like the source of the errors rather than the way to avoid them. As the procedures grow more complex, the likelihood of entering data accurately and consistently drops.

In almost every department within the business, the increased complexity of performing the job has meant more time required to do the same work. This means even less time and attention to data integrity issues and more dirty data - where someone may have taken the time to find out what an error message means and to address the data entry error prior to the implementation of ERP, now they may pass the error without addressing it because of the work backlog.

Out of necessity, though, where data integrity is compromised, user involvement into testing and reporting procedures is increasing. The correction will begin with an investigation by system/support and/or marketing services of the problem. Then, reporting tools will be generated to find all instances of the error. Finally, users will be enlisted to implement corrections. CSR involvement in the corrections is critical because of their intimacy with the data and as a training tool - those repeating the error most frequently will have the most corrections to make. The more corrections that the CSR is required to make, the greater the likelihood that they will be able to avoid the mistake in the future.

Counteracting and preventing dirty data - areas for improvement
A systematic approach should replace the more reactive crisis management approach to data integrity. Data audits should include daily integrity checks within the system and regular audits performed by user groups. Problems uncovered in those audits should be shared with all affected parties. The causes, effects, and resolutions of those problems should be systematically documented and stored so as to be easily accessible to interested users. If a similar problem occurs, documentation of other similar instances would be readily available. Where necessary, the communication should be followed up by training.

Regular training sessions should also be scheduled to ensure that users understand data integrity concepts and methods. These sessions would not only build a shared interpretation of data and preferred processing methods, but would also foster a more global perspective on the part of the users - instead of seeing only their own role, users would see their role in the context of the business. This perspective would assist in paring down some of the current procedural complexity. Simpler procedures would further increase data accuracy.

Equipped with an understanding of the impact of their work on other areas of the business, users can be analysts rather than data entry clerks. Analysts can make good decisions in a complex and dynamic work environment. Broadly-trained analysts would also be in a position to work effectively with consultants thus reducing our reliance on them.

Performance measures should be taken regularly to gauge the effectiveness of and to improve on the system and training initiatives. All of these measures would directly improve data integrity and would serve to underline the importance of data integrity to all users. These measures would reduce errors in carrying out tasks throughout the business and all their associated costs and help to draw a sharper picture of the business to improve long- and short-term decision-making.

6.0 Conclusion

Implementing ERP systems requires reinventing the business. Several legacy systems are integrated in the process with a single integrated system for managing operations across the organization. Data that resided in dozens of disparate sources is now available through one integrated system for all users in an organization. To achieve success in ERP systems implementation, project champions should make sure that they address the relevant issues. Some of the key lessons from this study include, among others, the following issues:

- The champion of the ERP implementation project should ensure that the transformation is not viewed as an IT initiative, rather a business necessity. This requires educating the stakeholders about the transition to an ERP.

- The champion for the change to ERP should recognize the value of data as an organizational resource and educate users about it. The issue of sharing corporate data
and assigning responsibilities for managing it should be done with a view to avoid any political issues arising from owners of disparate data sources.

- The ERP implementation should be planned to prepare users for the change. The expectations based on new responsibilities should be outlined upfront to avoid any conflicts.

- The user community should be given time to accept the changes in their work environment to minimize the impact on organizational culture, such as overcoming comments like "we've always done it this way". Users should be encouraged to use the new system by providing incentives.

- All data that is migrated to an ERP system should be cleaned before the migration. Automated tools for data migration can be very useful for companies.

- Training users on a continual basis is very important. It is important that users do not get bogged down by activities that take up too much of their time.

- Extensive testing is required for implementing ERP systems. A good strategy would be to phase-in the implementation rather than making a direct conversion.

- Consultants experienced with ERP implementation can be very helpful. Care must be taken to ensure that all the work done by consultants is documented for future use.

In this paper, we listed and discussed issues pertaining to ERP implementation. Though implementation in different organizations can vary based on the organizational culture and business needs we feel that the lessons from this case study would be valuable for organizations planning to implement ERP systems.

References


Reproduced with permission of the copyright owner. Further reproduction or distribution is prohibited without permission.