

Factors Influencing Error Recovery in Collections Databases: A Museum Case Study

Paul F. Marty
College of Information
Florida State University

marty@ci.fsu.edu

Abstract

This article offers an analysis of the process of error recovery as observed in the development and use of collections databases in a university museum. It presents results from a longitudinal case study of the development of collaborative systems and practices designed to reduce the number of errors found in the museum's databases as museum employees packed and moved their collections over a period of five years. Drawing upon a specific set of eighteen months worth of records documenting error rates and error management techniques, this article identifies three factors that influenced the ability of museum staff members to recover from errors found in their collections databases. The article concludes by examining the potential impact of these factors on the design of future collections databases in order to shed light on the wider issue of error recovery procedures in all information organizations.

Keywords

Museum informatics; error recovery; collaborative information systems.

Notice

This is a preprint of an article accepted for publication in *Library Quarterly*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. The final edited, revised, and typeset version of this paper was published in *Library Quarterly*, 75, (3), 295-328.
<http://www.journals.uchicago.edu/LQ/journal/contents/v75n3.html>

1.0 Introduction

Libraries, museums, archives and other information organizations depend heavily on the accuracy of their collections databases, yet anyone who regularly uses a library or museum catalog knows that these systems are not free from errors [1, 2]. This is perhaps to be expected: humans and machines both make mistakes, and the more complicated a database system, the more likely mistakes will occur [3]. What makes errors in collections databases such a nuisance, however, is that unexpected errors often lie in wait just beneath the surface, appearing all too frequently when one is not looking for them, yet tantalizingly difficult to detect and correct when one is [cf. 4, 5]. For librarians, museum professionals, archivists, and other information professionals working with databases containing thousands or millions of records vitally important for the day-to-day operation of their organizations, undetected errors can be a nightmare.

The risks posed by errors in collections databases present certain challenges to the developers and users of information systems in libraries, museums, archives and similar organizations. It can be extremely difficult to predict all the ways in which errors might occur, yet many collections databases employ data validation procedures that by definition rely on the designer's ability to predict potential errors [6]. When unexpected errors occur, they often slip past existing error prevention techniques to hide, undetected, somewhere in the system. Likewise, it is equally difficult to design information systems capable of helping users cope with unexpected errors, yet when errors are discovered, it is often by a surprised employee or patron who is in a position neither to correct the error nor to report it to someone who can [7]. Under these conditions, errors may be "discovered" then "forgotten," slipping unresolved back under the surface of the collections database, free to threaten another client or disrupt the work of another employee in the future.

To help developers and users cope with errors in collections databases, many studies have explored the relative advantages of different data quality control techniques for preventing, detecting, and correcting errors in information systems [8-10]. Despite hopes that these studies would identify universally-applicable methods for successfully preventing errors in collections databases, it turns out to be impossible to prevent all possible errors from occurring in any but the most artificial of environments. It is important, therefore, that designers of collections databases promote error recovery procedures that allow errors to be detected and corrected after they occur yet before they cause too much trouble for information professionals.

Studies of successful information systems document the need to encourage users to report errors when they find them, accept errors as a normal part of work processes, and learn from errors, thereby increasing overall system robustness [11, 12]. Designing databases that encourage such techniques, however, can be extremely difficult. While we know a great deal about the ways in which errors can occur in libraries, museums, and archives and have used that knowledge to correct many known errors in collections databases [5, 7, 13, 14], we still have much to learn about how the process of error recovery can be supported in the case of unexpected errors. Errors by their very nature are elusive, and we lack a suitable, conceptual framework that governs how information professionals can recover from errors found in collections databases.

In part this is due to the small number of detailed, longitudinal studies of libraries, archives, museums, and other information organizations where error recovery is encouraged, as well as the inherent difficulties of gathering data about the complex processes whereby error recovery occurs. One way to better our understanding of these processes may be study collections databases with fewer records than most information systems yet as many (if not more) problems with unexpected errors. Such databases would allow us to focus our efforts on a manageable set of data yet still provide a suitable number of instances of error recovery for analysis. A study of how information professionals cope with a large concentration of errors in a relatively small collections database, therefore, could have valuable implications for the design and development of future information systems for information organizations.

The study presented in this paper analyzes the process of error recovery as observed in a relatively small, yet extremely error-prone collections database belonging to a university museum. From 1998 to 2002, the author conducted a longitudinal, instrumental case study of a museum inventorying, packing, and shipping museum artifacts during a move from old to new facilities [15]. In this process, it proved impossible for museum employees to prevent 100% of all possible errors from occurring. Over time, however, museum staff members learned to live with errors, developing new mechanisms which allowed them to recover from unexpected mistakes. As a result, they were able to correct nearly 100% of all errors which occurred during the move, before the move was completed. Moreover, they discovered that by emphasizing error recovery over error prevention, they were better able to develop more robust collections databases over the time period of the study [12].

This article draws upon this analysis to present an overview of one particular portion of the museum's error recovery processes: those which supported error detection and correction in the collections databases which governed the museum's packing process. It discusses the error recovery procedures implemented by museum staff members over a period of eighteen months, and provides examples of the types of errors the museum employees were unable to prevent and how they recovered from those errors by detecting and correcting them before they were able to cause irrevocable damage. Based on a qualitative analysis of over six hundred error transaction logs documenting the error recovery procedures in use at the museum, the paper identifies three factors responsible for encouraging error recovery procedures in collections databases. The paper concludes by discussing the potential impact of these factors on the ability of information professionals to promote error recovery practices in their own information systems and the benefits of this approach for libraries, museums, archives, and other information organizations in general.

2.0 Literature Review: Error Management Techniques for Information Systems

Databases in all information organizations suffer from dangerously-high error rates [6, 10, 16, 17]. According to Jacsó [18], there is an "appalling amount of sloppiness in bibliographic and numeric databases" (p.38). These errors can be devastating for both the producers and users of information systems [19]. The costs of database errors for commercial organizations, where typical error rates per field range from one to five percent, can be enormous [20]. Errors in the collections databases of libraries and other information organizations can be equally disastrous, and have even been the cause of death in at least one case [4].

Many have documented the types and quantities of errors found in databases in libraries and other information organizations, and the results are not encouraging [5, 7, 14, 21]. There is also an extensive literature on data quality, where various researchers have advocated different methods for improving the quality of data in information systems [22-24]. These researchers have attempted to clarify the relevant issues by defining different dimensions of data quality [9, 25], exploring the metadata requirements for a focus on data quality [26], or determining data quality enhancement priorities based on cost-benefit analyses [27].

Any attempt to provide even a cursory survey of this research literature in its entirety would be far beyond the scope of this article [cf. 1, 2]. We therefore limit ourselves to a few key points that pertain to attempts to manage errors in information systems or collections databases. We will begin by examining two different approaches to error management: preventing errors before they occur; and detecting and correcting errors after they occur.

2.1 Designing to Prevent Errors in Information Systems

There have been many attempts to prevent errors from occurring in information systems, most focusing on data validation procedures built into information systems that check data quality attributes at the time of data entry [6]. Such validation procedures, however, are not 100% successful, and even data once thought to be correct can become recognized as erroneous at a later time. Errors, it seems, are the inevitable result of “the processes of data production and data utilization” [3, p.99; cf. 28]. Why is this the case?

Research into collaborative workflow systems, where multiple individuals use, access, or manipulate database records at different locations and different times, can shed light on the inevitability of errors and why it is so hard to prevent errors from occurring in information systems. Efforts to prevent errors in these systems tend to enforce rigid and inflexible models of work that are incapable of adapting to meet unforeseen contingencies [29-31]. While this may appear in theory to be a good way of preventing errors, in practice it is nearly impossible for users and designers to specify all cases, negotiate complex rules, or articulate tacit knowledge in advance [32-37]. The inherent complexities of life and the changing nature of work require that most activities adapt to unforeseen events and errors as inevitable aspects of human performance [38-41]. Information systems that attempt to prevent errors by strictly controlling what people can and cannot do tend to fail because they are insensitive to the changing needs of their users and lack the flexibility or capability to cope with errors and unexpected events [42-44].

These studies have prompted a change in mindset where researchers attempt to develop information systems that support all types of work activities, including errors and unexpected events, instead of supporting only known work activities. From this perspective, errors are no longer unwanted events which must be controlled by overly rigid systems; they are things to be planned for rather than avoided at all costs. Instead of designing *to prevent errors*, therefore, one designs *for errors*. In this article, we will refer to this distinction as error prevention vs. error recovery, and we will now examine attempts to encourage error recovery in information systems.

2.2 Designing for Error Recovery in Information Systems

Several recent streams of research have focused on the topic of designing for errors when building information systems [3, 46, 47]. Many researchers have studied how users deal with the inflexibility of information systems or how designers build information systems capable of coping with unexpected events [48-50]. These studies share the common finding that in order to design for error recovery, developers need to build systems that involve users in the process of looking for errors, and encourage them to play a direct role in improving data quality.

This idea has not always been well received by researchers or system designers: “in general, the data quality literature argues that users are not very capable in finding errors in data” [24, p. 28]. Several studies have demonstrated that users are unlikely to discover errors as they work with information systems, even if the data they are using are of direct importance to them [16, 51, 52]. More recent evidence, however, indicates that users can play an important role in detecting and correcting errors if properly encouraged and empowered to do so [6]. It may be that users are better at detecting errors in data than first assumed, but that most information systems do not provide users with the ability to correct errors when they find them. How, then, can we better enable users to detect and correct errors in data while working with information systems?

To answer this question, researchers have explored methods of helping users document, track, and report errors in systems or encouraging users to become more directly involved in the error recovery process, allowing them to correct errors on the fly as they discover errors during their normal work. In a Finnish paper mill, for example, an electronic diary provided a means of communicating problems and potential resolutions between workers, allowing error diagnoses and repairs to occur as part of the ongoing workflow [45, 53]. Others have suggested that information systems incorporate a digital history of data modifications that tracks not only all previous values but also why each value was changed, when, by whom, and with what authority [54-56]. Similarly, Robbin and Frost-Kumpf [3] document the need for simple communication mechanisms for reporting and tracking errors: they suggest the creation of a “permanent repository of conversations about error” that would incorporate the expertise of multiple users and create a trail to follow when tracking errors; they also suggest the need for a “supported learning environment to encourage the self-detection of error” (p. 116). Provided there are appropriate and easy-to-use feedback mechanisms for reporting errors, such techniques can lead to overall improvements in data quality [57].

2.3 Detecting and Correcting Errors in Collections Databases

There have been many similar attempts to encourage users to detect and correct database errors in libraries, museums, archives, and other information organizations. Inspired by the high number of errors in collections databases, researchers have explored different ways to help users seek out and correct errors. These analyses have focused primarily on methods of assessing the quality of information systems, determinations of the types of errors most databases contain, and studies of how errors propagate from system to system [1, 2, 5]. Evaluation methods such as the “Dirty Database Test” [58, 59] have primarily focused on finding (and hopefully correcting) typographical errors in databases [14, 60, 61]. Although most agree that databases contain a large number of errors [21], there has been some disagreement over whether these errors need to be fixed [62] and whom should be empowered to do so [7]. Noting the impossibility of achieving

100% accuracy, some researchers advocate correcting errors in those records likely to be most damaging to the end user [13, 63].

There has been some discussion about the potential for employing end user feedback to improve data quality in collections databases; the methods suggested, however, are often laborious, thereby increasing the effort required to volunteer information and reducing the motivation to contribute [64]. Davis [65], for example, describes an attempt to encourage Online Union Catalog users to report errors which had an extremely low response rate, in part because of the need to supply supporting documentation. Similar difficulties have led others to propose a variety of solutions for encouraging user feedback, including fee-based ones: in return for reporting errors, online searchers would be credited for the cost of accessing the erroneous data [63]. Others have advocated the creation of online bulletin boards to provide a forum for searchers to complain about database errors and for vendors to respond [66]. Common to all of these approaches is the need to overcome the inherent difficulty of encouraging users to participate in the process of detecting, correcting, and reporting errors.

2.4 Seeking a Conceptual Framework for the Error Recovery Process in Collections Databases

One reason for this difficulty may be the lack of a conceptual framework that explains the process by which errors can be detected and corrected by users in information organizations. We have a very hard time understanding what motivates individuals to participate in the error recovery process or how information systems can best encourage this participation. Without this understanding, it is extremely difficult to develop collections databases that encourage information professionals to take an active role in correcting errors as they are discovered. According to Adams [4], “How can information science deal efficiently with correcting information in this imperfect world? What mechanisms do we have on hand that can help us cope with errors and control or prevent their perpetuation? The simple answer [...] seems to be, ‘Not many.’” (p. 21).

To help information professionals cope with errors, we need a model that explores how people working with collections databases can be encouraged to play a role in recovering from errors found in those systems. We do not have a good understanding of the characteristics of collections databases systems that encourage their users to detect, correct, and recover from errors vs. those that force users to ignore or workaroud errors each time they are encountered. Without this understanding, we will be unable to design and develop collections databases that can help information professionals react to unexpected errors positively, correcting or flagging errors for future correction as they are discovered. There is a need, therefore, for more studies, especially long-term, longitudinal case studies, that explore different mechanisms of encouraging error recovery as a part of the day to day operations of any given information organization.

This article, therefore, will examine the process of error recovery in a university museum which used relatively straightforward, yet extremely error-prone collections databases to govern the packing and shipping of its artifacts. We will explore the following research questions: What are some of the factors that influence the ability of users of collections databases to recover from errors? How can users and developers of such systems encourage the error recovery process in a way that improves the overall robustness of information systems? Answering these questions

will take us one step closer to having a better understanding of the error recovery process in action. Although we will remain far from the ultimate goal of developing a conceptual framework to explain this process in all information organizations, the results of this study will shed light on the broader issue of encouraging error recovery in general.

3.0 Research Methods: Studying Error Recovery in a University Museum

The Spurlock Museum is the museum of world cultures at the University of Illinois at Urbana-Champaign. From 1997 to 2002, staff members at this museum worked to move 45,000 artifacts across campus from old to new facilities. These collections had previously been stored and displayed in the attic of a 100-year-old building, with little space, no climate control, and poor lighting. In March 2000, construction was completed on a brand-new modern museum facility, with nearly three times the square footage of the previous facility. The museum staff had only five years to inventory, pack, and move their artifacts, while simultaneously working with exhibit designers and architects to design, build, and install all of the exhibits in the new facility.

Museum staff members had never done anything like this before, and they had in place neither the systems nor procedures necessary to support this move. They had to train employees to inventory and digitize information about their artifacts (the museum's only previous attempt at a full inventory had been in 1973). They had to develop methods for packing their collections so that they could be safely moved across campus. They had to develop procedures to track and move thousands of boxes to the new facility without losing a single artifact. To support all of these activities, they needed to design and develop collections databases and other information systems from scratch in a very short period of time [67-68].

The development and use of these collections databases provide valuable data that can be used to inform a study of error recovery practices. The information systems developed by museum staff members to coordinate the process of moving the museum were extremely susceptible to error. These systems had to support the collaborative practices of the many different stakeholders working in different areas of the museum (registration, collections, exhibit design, etc.) as they worked together to move the museum. Since museum staff members had never done anything like this before, they faced the potential of an extremely high error rate as they developed their collaborative systems and practices over time.

From 1998 to 2002, the author conducted a longitudinal, instrumental case study [69, 70] to study the evolution of the museum's information systems as they coped with these challenges [15]. During this time, the author worked at the Spurlock Museum as Director of Information Technology, and served as a participant observer in the error management process, conducting ethnographic observations of error recovery procedures implemented by museum employees [71, 72]. For these reasons, the author had access to a wide variety of data, including manuals that document the practices of the museum staff, copies of the museum's collaborative systems, and hundreds of error transaction logs that show how museum employees detected and corrected the errors they found their systems at all stages over a period of five years. The availability of these data offers a unique opportunity to shed light on the process of error recovery in action.

3.1 Drawing upon Multiple Approaches to Study Error Recovery Procedures

The primary theoretical basis for this study was grounded theory informed by a number of additional theories of how data quality is controlled in error-prone situations [73]. Recently, several researchers have argued for the use of multiple theories and approaches to studying error management in information organizations, and many different theoretical frameworks have been proposed in the literature to evaluate error management procedures. Klein [6], for example, has used Theories of Individual Task Performance [74] to study aspects of user behavior that influence error recovery practices. Robbin and Frost-Kumpf [3] argue the need to apply social science theories to error analysis studies, demonstrating how theories from organizational science, cognitive psychology, and related areas can benefit analyses of error management processes [e.g. 75-78].

While each theoretical framework has its own advantages and disadvantages, there are two things they all share in common. First, each takes into consideration the fact that no attempt to prevent errors will be 100% successful; errors are essentially inevitable in all information system. Second, each offers a similar understanding that for error recovery procedures to be successful, the users of information systems must be empowered to play a role in correcting errors found in those systems; this hold true whether these users are actively seeking errors or whether they stumble across errors in their daily work. Thus, to understand the error recovery process completely, we need to find and analyze those situations where unexpected errors in collections databases are discovered and where users are empowered to deal with errors at those times.

Finding these examples of error recovery in action can be very difficult because most error recovery activities (especially those which were unsuccessful) are not typically recorded or discoverable after the fact. The relatively straightforward nature of the collections databases used by museum staff members, however, combined with their habit of meticulously recording information about the process of moving the museum as part of their error management practices, helps bring these examples to the forefront in this study. The available data clearly document both the errors faced by museum staff members as they moved the museum as well as the procedures they implemented to try to prevent, detect, correct, and recover from those errors over time. By looking at all the ways that museum staff members coped with errors in their collections databases, we can identify factors that encouraged museum staff members to recover from the errors which inevitably occurred as they moved their museum.

3.2 Looking for Errors in a Museum's Collections Database

In order to reduce the data set to a more manageable level, this study examines the error recovery process as observed in the development and use of one specific set of collections databases: those which supported the packing of the museum's collections. From August 1998 to March 2000, museum staff members packed 27,166 artifacts into 1408 boxes, introducing many errors into their collections databases as a result. The records created by museum employees as they resolved these errors while packing their collections document in detail the different types of errors they faced and how they dealt with them.

The available data provide us with an opportunity to study the error recovery process at the Spurlock Museum in two different ways. First, by examining hundreds of detailed error transaction logs which document the source and resolution of each error found by museum staff members, we can study those places where errors occurred during the packing process and how museum staff members resolved those errors, including the types, numbers, and frequency of errors. Second, by examining dozens of changes to the museum's procedures and manuals, we can study how museum staff members attempted to reduce the number of errors through a variety of error management techniques, including the impact of these methods on their ability to recover from errors and their consequences.

Having identified those instances where museum staff members attempted to cope with errors in their collections databases, therefore, we took a grounded approach to analyzing the processes by which museum staff members attempted to recover from errors and improve the quality of data in their information systems. Analyzing the errors that occurred in the packing of each of 1408 boxes, we used a process of coding and memoing, as outlined by Strauss and Corbin [73], to develop an understanding of the museum staff members' activities and to create a list of dimensions that influenced the error recovery process in the museum. This kind of analysis was an iterative, on-going process where records documenting the error recovery process from each packed box were analyzed as part of a continual process of exploring the data to identify emergent themes and dimensions. By studying how museum staff members attempted to prevent, detect, correct, and recover from errors in the packing process, we identified several factors that influence the ability of museum staff members to recover from errors in their collections databases. The result of this analysis, therefore, will provide us with a better understanding of how information professionals working in any information organization can recover from errors in their own information systems.

4.0 Results: Coping with Errors in a University Museum

The packing process at the Spurlock Museum was a collaborative workflow system that coordinated the efforts of ten full-time staff members and several dozen part-time students and volunteers working to move the museum's collections [15]. Museum staff had no prior experience with projects anywhere near this size or of this complexity, and the large number of individuals collaborating on this project with various levels of expertise meant that there was great potential for errors to occur. Employees might mistakenly misread an inventory number when packing a box, for instance, thereby causing another employee to record in the museum's collections database that the wrong artifact was packed in the wrong box. Making matters worse, this kind of error might daisy-chain from box to box, eventually leading to great confusion as to where any given artifact actually was. If uncorrected, these artifacts would be essentially "lost in the system," and pose problems that could disrupt the entire move to the new facility.

4.1 Error Transaction Logs: Analyzing Errors in the Packing Process

To keep errors under control, museum staff members decided early in the project to record as much data as possible about the packing process [12]. One of the things they tracked very carefully was each incident of error, noting in each case what the problem was, how it occurred,

and how it was resolved (see example below). The lifecycle of an error, therefore, from its discovery to its resolution, was completely recorded in an “error transaction log.”

Error transaction logs served not only to document errors but also as the principal medium of communication between the two museum departments primarily responsible for packing the museum: Registration and Collections. Museum staff members working in the Collections department packed artifacts into boxes, and in the process completed a paper form called a “Packing Data Sheet” which recorded extensive data about the packed box, including which artifacts were packed in that box (for more details about the packing process, see [15]). Employees working in the Registration department then entered data from the Packing Data Sheets into the museum’s collections databases; most data were stored in the “Packing Database,” which tracked the location of each box, as well as the artifacts contained within, throughout the move.

Errors introduced as artifacts were packed were usually discovered at the moment when information from a Packing Data Sheet was entered into the Packing Database. If an error was found, this was noted on the Packing Data Sheet, and the form returned to Collections as a “problem form.” A Collections staff member, usually the Assistant Collections Manager, retrieved these problem forms on a regular basis and attempted to resolve the listed problems. The process of solving each problems, as well as its eventual solution, was recorded on error transaction logs, and returned to Registration along with the corrected Packing Data Sheet to be re-entered into the Packing Database. By analyzing these error transaction logs, we can explore the sources and types of errors that occurred, the ways in which these errors manifested themselves, and the frequency in which these errors occurred at the museum.

Depending on the type of error, error transaction logs took several different formats. In the simplest of cases, errors were resolved by simply writing the solution directly on the Packing Data Sheet, so that the error transaction log was actually contained in the margins of the form. Usually, however, error transaction logs were written on sticky notes which were then attached to the Packing Data Sheets directly. In the most complex cases, error transaction logs were typed as separate documents in a word processor, then printed and attached to the Packing Data Sheet. Errors frequently involved multiple boxes, however, and over time, individual error transaction logs typically became covered with additional layers of annotations and sticky notes recording information about related problems. In each case, a summary of the error transaction log was also entered into a comments field in the Packing Database.

Packing problems were typically the result of simple transcription errors, where artifact accession numbers were either miswritten onto Packing Data Sheets or misread when entered into the Packing Database. There were many different reasons why such errors might occur. Inventory numbers were often very difficult to read on the artifact itself: numbers may have deteriorated over time and were often written on corners of the artifact in crabbed handwriting. Some numbers were written ambiguously (e.g., the number “18.1.18” could be read equally well upside-down as “81.1.81”, or unclear letters could be mixed in with the numbers, making an “8b” look like an “86”). In addition, different objects frequently used different numbering schemes, and museum employees often had problems keeping straight which numbers used

which format. From such simple mistakes, errors then infiltrated the museum's collections databases in many different ways.

To manage different types of errors, museum staff members evolved their own terminology for describing errors. The two most common types of error were called "no such number" and "already packed" errors. The former type occurred when an artifact was packed that did not exist in the museum's collections database; while this could indicate the presence of an uninventoried artifact, the more typical explanation was that a number was miswritten on the Packing Data Sheet or mistyped into the Packing Database. The latter type occurred when an artifact was packed that, according to the museum's collections database, was already packed in another box; although this error could be the result of a simple typographical mistake, more frequently it indicated the discovery of separate, previously unknown error (see below). A subtype of the "already packed" error occurred when the same accession number was recorded more than once on the same Packing Data Sheet; these were called "duplicate number" errors, and also typically indicated more serious problems where an artifact might be missing from the box inventory.

While many of these errors had the potential to be caught and fixed at the time of data entry into the museum's collections databases, staff members also faced more serious errors that could not be caught at the time of data entry, and remained hidden until a subsequent problem (such as an "already packed" error) uncovered them. These were called "latent" errors, and their resolution typically required a bit of detective work on the part of museum staff members. Consider, for example, the situation where an artifact is packed into a box by an employee who miswrote the object's accession number on the box's inventory. Usually, this error would be uncovered as a "no such number" problem. If by chance, however, the miswritten number was associated with a different artifact that was not yet packed, no error would be detected and that artifact's record would be marked erroneously in the collections database as being packed in the wrong box. At some future point, when the misidentified artifact was actually packed, an "already packed" problem would occur, and museum staff would have to follow a trail of mistakes back to the original box to uncover the latent error. The following example shows how this situation might occur, and in the process illustrates the main types of errors faced by museum staff members.

Error Recovery Example 1: Box 1340

On January 25, 2000, thirty-three artifacts were packed into Box 1340 and recorded on a Packing Data Sheet (see Figure 1). When a Registration employee entered this form into the Packing Database on January 28, 2000, she found several errors. The first accession number listed under "Box Contents" did not correspond to any object in the museum's collections database; she therefore circled the number and wrote "accession number not found in database" next to it. Next, this employee observed that, according to the museum's collections database, artifact number 1922.01.0011 was already packed in Box 531; she circled the potentially erroneous number and wrote "Database has this in Box 0531." Finally, the Registration employee noticed that accession number 1954.04.0013 was written on the Packing Data Sheet twice; she therefore circled both duplicate numbers, connected them with a line, and wrote the word "Duplicate" on the form. Thus, all three of the main types of errors occurred on this one Packing Data Sheet.

After the Packing Data Sheet was returned to Collections, the Assistant Collections Manager opened the box and double-checked its inventory on February 16. She confirmed that the first accession number was indeed wrong and should have been 1921.01.0007. Next, she discovered that there was no artifact with accession number 1922.01.0011 in the box and that the error was caused by a simple transposition mistake: the correct number was 1922.10.0011. Finally, she confirmed that the duplicate number error was simply written on the form twice and no other object was lurking unrecorded in the box. The total number of objects in the box therefore decreased by one, from 33 to 32. The Assistant Collections Manager then returned this Packing Data Sheet to Registration for re-entry into the Packing Database.

In Registration, the Packing Data Sheet remained unprocessed until April 26, 2000, when the museum's Registrar made the changes to the database according to the Assistant Collection Manager's solutions. However, she found that she could not correct the change from 1922.01.0011 to 1922.10.0011 because, according to the collections database, that object was already packed in Box 1403. Affixing a sticky note with the comment "1922.10.0011 is in Box 1403," she returned the Packing Data Sheet to Collections. Later that same day, back in Collections, museum staff members reconfirmed that 1922.10.0011 was indeed packed in Box 1340, and an employee was assigned the task of opening Box 1403 to sort out this new problem. When she did so, she found that Box 1403 was riddled with its own problems; soon, museum staff members had discovered four latent errors which had been hiding in Box 1403 for nearly three months which they now needed to solve.

[INSERT FIGURE 1]

Given such a wide variety of potential problems, museum staff members expected to make mistakes; the number of mistakes actually made, however, surpassed what anyone had imagined. Out of 1408 boxes packed, 373 of them had inaccurate inventories, an error rate of 26.5%. These errors affected a total of 627 artifacts out of the 27,166 artifacts packed during the time period of the study, an error rate of 2.3%. Chart 1 depicts the number of problem boxes occurring out of the number of total boxes packed each day.

[INSERT CHART 1 HERE]

The number of "problem artifacts" found in any particular "problem box" ranged from one to ten, with 241 boxes having one problem artifact, 69 having two problem artifacts, 36 having three problem artifacts, and the remaining 27 boxes having anywhere from four to ten problem artifacts. Whether a box contained one problem artifact or ten problem artifacts, however, what mattered to the museum staff members was that they had to re-open, double-check, and re-pack an already packed box. This analysis, therefore, intentionally focuses on the "box level" instead of the "artifact level" because that approach more accurately reflects the focus of museum staff members involved with detecting and correcting errors in the packing process.

From that perspective, it is important that we examine the types of errors discovered in each box, how they occurred, and when they were detected. The majority of errors (affecting 489 artifacts)

were detected during the data entry process; these included 210 “no such number” errors, 221 “already packed” errors, and 58 “duplicate number” errors. The remaining 138 errors were the more serious “latent” errors, which could not be so easily detected, but instead had to be recovered by museum staff members after the fact. Each of these errors generated error transaction logs that document the error recovery procedures in use at the museum. See Tables 1 and 2 for a summary of these figures.

[INSERT TABLE 1 HERE]

[INSERT TABLE 2 HERE]

4.2 Controlling Errors: Procedural Changes and their Impact on Error Rates

Despite facing such a high rate of errors, museum staff members corrected more than 99% of these errors before the completion of the move. They accomplished this by developing excellent error recovery procedures that assisted them in the process of detecting and correcting the vast majority of their errors. The way this was accomplished provides an excellent opportunity to analyze the factors that encourage error recovery in information organizations. We turn now, therefore, to an analysis of the procedural changes implemented by museum staff members to try to control errors, and the impact of those changes on error rates in the museum.

Museum staff members knew they were working on a very difficult and error-prone project. Carefully tracking errors and solving packing problems was a necessary burden if museum staff members were to live up to their own expectations for accuracy. Not even the simplest error could be ignored, as even simple errors could turn into more complicated problems, especially as errors daisy-chained from box to box.

Museum staff members also knew, however, that the majority of their errors resulted from the work of undergraduate students hired by the museum to assist in the move; Chart 1 (above) shows that the fewest errors occurred during the winter breaks, when very few undergraduate students were working in the museum. Many museum professionals would recoil at the thought of entrusting their collections, in either physical or digital form, to the care of undergraduate students with little experience in museum studies. Staff members at the Spurlock Museum, however, had to rely on undergraduate employees if they wanted to move the museum. While these students rose to the challenge and accomplished a great deal in a few years, senior museum staff members nevertheless recognized that students workers were a significant potential source of error and made some changes to help reduce the possibility of error.

In this section, we will look at procedural changes designed to reduce the likelihood of errors, the results of those attempts, and the consequent changes in approach to managing errors found in the museum’s collections databases. The following two examples of procedural changes represent attempts by museum staff members to prevent errors from occurring at the moment when the Packing Data Sheets were completed.

Procedural Change 1: Controlled Data Formats

One early change to packing procedures prohibited undergraduate students from using abbreviated artifact accession numbers, requiring them instead to record numbers using their full numeric format. Although accession numbers at the Spurlock Museum officially had a tripartite format such as 1970.01.0021, students (and other employees) were accustomed to simplifying this format to numbers such as 70.1.21. Such simplifications, however, were open to misinterpretation, especially (as often happened) when students read numbers aloud to each other as they packed artifacts: what one said was not always what the other heard. For example, one might say the numbers “seventy... one... twenty-one” only to have the other hear, and write down, “seventy-one... twenty... one.” In this way, artifact 1970.01.0021 might be packed as artifact 1971.20.0001, leading to a packing error. Fully-formatted accession numbers, where everyone knew that numbers would follow a four.two.four digit pattern, were less likely to be misinterpreted.

Procedural Change 2: “Magic Fairy”¹

A procedural change implemented to solve one problem had the potential to reduce inaccuracies caused by misread numbers. Student workers assigned to pack the collections would often put off packing artifacts they perceived as being more difficult to pack or more fragile to handle; such artifacts would sit for weeks while other artifacts were packed almost immediately. To solve this problem, the Collections Manager instituted a new procedure where she and her assistant, the two museum employees most skilled in handling artifacts, would stay late every night and “pre-wrap” a day’s worth of artifacts to be packed in white ethafoam (the first step in packing any object). When the students returned to work the next morning, they were told that a “magic fairy” had visited the museum during the night and left pre-wrapped artifacts for them to pack. These pre-wrapped objects turned out to be much less intimidating for the students to pack. So that the students would know the accession number for each object, the Collections Manager would write the artifact’s identification number on the outside of the pre-wrapped piece, clearly and carefully in black ink on yellow tape. Thus, with one change, they not only made the task of packing artifacts easier on the students but they also made the task of reading accession numbers much simpler by introducing expertise (and clearer handwriting) into a crucial phase of the packing process.

Surprisingly, however, procedural changes that emphasized error prevention techniques such as these had very little effect on the overall error rate in the museum. No matter what changes museum staff members introduced into their packing procedures, they could not keep errors from occurring as boxes were packed.² Chart 2 shows the error rate of packed boxes with incorrect

¹ The phrase “magic fairy” reflects the actual terminology used by museum staff members to describe this procedure. While some might think the use of this phrase is inappropriate, it is very important that qualitative researchers maintain the “in vivo” terms used by research participants if at all possible [73, p.105].

² Specifying an exact date for when these procedural changes occurred is problematic. As discussed in detail in an article forthcoming in the *Journal of Computer-Mediated Communication* [79], changes to the museum’s packing practices were not always temporally clear-cut; instead, they evolved in a continual process of co-adaptation. When this is combined with the fact that the process of detecting and correcting errors in just one box involved many different individuals over a lengthy period of time, it becomes very difficult to establish a definitive date when each procedural change was implemented museum-wide. As discussed here, however, the important point is not when

inventories expressed as a percentage of the total number of boxes packed per day. The trend line shows that, with the exception of the two winter breaks, this error rate remained relatively constant at 26.5% over the course of the eighteen months. No matter how hard museum staff tried to improve the workings of the packing process, the error rate did not decrease.

[INSERT CHART 2 HERE]

Realizing that they were unable to prevent errors from occurring in advance, museum staff members therefore began to focus more on procedural changes that emphasized error recovery techniques by helping museum employees detect and correct errors after they occurred. As museum staff members learned what problems were likely occur, they could feed their knowledge back into the packing process, thereby reducing the likelihood of generating new problems and making the whole process more robust. The following two examples illustrate attempts to work within the system by taking advantage of known errors to improve the overall robustness of the museum's collections database.

Procedural Change 3: Contents Inventory

Museum staff members quickly learned that in solving one problem, they might, if they kept their eyes open, discover another problem which could be solved immediately. If so, they could offer both solutions simultaneously, removing an entire cycle from the problem solving process. For example, whenever an employee opened a box to solve a known problem, she could look beyond the immediate problem by taking the initiative to double-check the numbers of every object in the box once she had it open. This procedure was called a "contents inventory," and while performing a contents inventory was not mandatory, if a box happened to be opened to solve one problem, museum staff members frequently felt it was worth the effort to perform a contents inventory on the chance that they would discover another error.

Procedural Change 4: Pre-emptive Strikes

Similarly, museum staff members quickly learned that by using the museum's collections databases to "look several moves ahead," they could determine whether in solving one problem they were causing another problem somewhere else. If so, they could solve hidden packing problems before others found them, thereby saving the museum time and money. For example, whenever an employee corrected a known error, she could double-check the museum's collections database to see if by correcting that error, another error would be uncovered. She could then find the solution to that, as yet undiscovered, error, and provide all of this information in solving the first problem, in essence saying: "by the way, when you solve this problem with Box X, you'll find another problem with Box Y, and I'm giving you the solution to that one in advance."

these changes occurred but that their overall impact on the museum's ability to prevent errors from occurring in packed boxes was negligible.

These examples demonstrate the growing familiarity of the museum staff with the problem solving process in particular and the packing process in general. The more problems they solved, the better they became at detecting and correcting errors after the fact. As museum staff members tracked down and solved one packing problem after another, they were learning to live and work in a state of near constant error. Instead of increasing their ability to prevent problems, they improved their ability to catch errors, and recover from them. The following example shows how this familiarity paid off in a complex process of error recovery.

Error Recovery Example 2: Boxes 531/462/506

In December 1999, the museum's Registrar accidentally discovered that two different artifacts, an Egyptian Bottle Stopper and an Apulian Kantharos, shared the accession number: 1926.02.0254. Unfortunately, both of these artifacts had already been packed into two separate boxes: Boxes 531 and 462. Given the error detecting techniques built into the packing process, this was not supposed to be possible: one artifact should have registered as an "already packed" error. That this had not happened indicated a potentially serious problem with the museum's collections database. The Registrar therefore asked the Assistant Collections Manager to open Boxes 531 and 462 and determine if both artifacts were actually marked with the same accession number.

The Assistant Collections Manager first checked Box 531 and confirmed that the Egyptian Bottle Stopper was packed in that box and marked with that number. When she opened Box 462, however, she found neither an Apulian Kantharos nor any artifact with that accession number! Worried, she informed the Registrar of the missing artifact, but with box 462 open, she proceeded to perform a "contents inventory" on the box. As a result, she discovered an additional, previously undetected error: a small Roman Marble Fragment (1922.10.0049) had been mis-recorded on Packing Data Sheet as 1922.01.0049.

Before reporting this new error, however, the Assistant Collections Manager first performed a "pre-emptive strike" to ensure that in solving this error she was not creating another one. When she checked the Packing Database, she discovered that this Roman Marble Fragment (1922.10.0049) was already packed in another box: Box 506. According to the Packing Data Sheet for Box 506, however, no artifact with that number was packed in that box. Instead, Box 506 contained an artifact with a very similar accession number (1922.01.0049). She theorized, therefore, that when a Registration employee had entered this information into the Packing Database, he or she had accidentally transposed the two middle numbers. Since this transposition had produced a valid accession number, it was not detected as a "no such number" error. As a result, artifact number 1922.10.0049 ended up being "packed" in two boxes (462 and 506), yet this error was not detected as an "already packed" problem because the two separate transposition errors had cancelled each other out.

To confirm this explanation, the Assistant Collections Manager opened Box 506 to double-check the identity of the object with the accession number 1922.10.0049. To her surprise, that artifact turned out to be the missing Apulian Kantharos. By tracing this

error through database backups, museum staff members were able to discover what had gone wrong. An error early in the inventory process had led to the creation of two records for the Apulian Kantharos in the collections database, each with a different accession number. Through additional errors, these two “records” were each packed into different boxes, and later confused with the Egyptian Bottle Stopper and Roman Marble Fragment respectively. If it had not been for the museum staff’s emphasis on error recovery, the Apulian Kantharos would likely have remained missing until after the move.

The almost daily activity of tracking down and correcting errors such as these gave museum staff members an improved understanding of how their information systems worked overall. One way of measuring museum staff members’ improved ability to understand and recover from errors in their collections databases was to see how many errors escaped their notice. Although they managed to catch 627 problem artifacts as they packed the collection of 27,166 objects, were these simply the tip of the iceberg? Were there another 600 errors waiting to be discovered that they had missed? The answers to these questions were determined by studying the unpacking process: while preparing the museum to open to the public, museum staff members unpacked their boxes, placing their artifacts into storage or preparing them to go on display. As each box was unpacked, a box inventory was completed that could be compared against the museum’s collections databases. By comparing these new inventories with existing information about which artifacts were supposedly packed in each box, we discovered that only four artifacts, out of 27,166, were problems that could have been caught during packing but were not. Therefore, nearly 100% of every single problem that could have been caught before the move was in fact caught and corrected by museum staff members (see Table 3).

[INSERT TABLE 3 HERE]

4.3 Living with Error: The Overall Benefits of the Error Recovery Approach

By changing their focus from error prevention to error recovery, museum staff members gained many benefits that went beyond their ability to detect and correct errors. Tracking down and correcting even the simplest error helped them improve and evolve their information systems overall [12]. These results echo the findings of other studies which illustrate how the process of correcting errors can have benefits that go far beyond the immediate benefits of fixing errors [11; cf. 3]. By developing procedures to recover from errors in their collections databases as they were discovered, Spurlock staff members were able to more than simply correct errors: they learned more about how their information systems functioned and were able to use this knowledge to improve the process of packing the museum. An analysis of the factors that encouraged these error management techniques, therefore, could help information professionals gain insights into the benefits error recovery could have for their own collections databases.

The examples discussed above illustrate certain characteristics of the collections databases in use at the museum that emphasized error detection and correction and improved the error recovery process. In the first example (Box 1340), we saw how museum staff members needed to work within the system to communicate between different museum departments and collaboratively resolve errors. In doing so, they relied on the system’s ability to track changes (both on paper and in the database) to backtrack packing problems and determine potential sources of errors. In

the last example (Boxes 531/462/506), we saw how a familiarity with the collections database enabled museum employees to solve previously undetected problems and to cope with situations where seemingly unrelated errors interacted in unexpected ways. In all of these situations, museum staff members needed to work with errors in their information systems, using characteristics of their collections databases that encouraged error recovery to their advantage.

Based on these and more than six hundred other examples of error recovery analyzed during this study, we developed three factors we believe account for the museum staff's ability to recover from errors. These factors, along with their implications for the design and use of collections databases in general, are now discussed in detail.

5.0 Discussion: Factors Influencing Error Recovery in Collections Databases

From our analysis of the way staff members at the Spurlock Museum coped with errors as they packed their museum, we developed the following three factors which we believe play an important role in influencing error recovery processes in information organizations. Each factor, we argue, describes a different aspect of those environments most likely to support error recovery activities and result in an overall reduction of errors in collections databases. In the museum, these factors were a natural outgrowth of the information systems (including the Packing Database and Packing Data Sheets) used to pack the museum; they illustrate aspects of the museum environment which encouraged error recovery. If these factors could be enabled in collections databases in other information organizations, we believe they could help designers create information systems capable of encouraging error recovery at the hands of users.

5.1 Error recovery is encouraged when collections databases are easy to understand and manipulate, helping users feel more comfortable with the system and allowing them to more easily recognize and attempt to resolve errors.

As illustrated by the examples above, the information systems used by museum staff members to pack their museum were relatively easy for them to understand and manipulate. Not only were the museum's collections databases relatively easy-to-learn, but the overall packing process involved elements which were physically manipulated: paper forms moved from department to department, artifacts moved from table to table, boxes moved from room to room. The openness of these activities contributed a visible, tangible aspect to the process: every employee, from undergraduate students to senior staff, could observe the packing process in its entirety. As a result, museum staff members quickly learned how the components of their information systems functioned, as well as how each could be manipulated to give them a better chance of solving packing problems.

This high level of familiarity with the museum's information systems played a significant role in the museum staff's ability to detect and correct errors in their databases. First, it helped them recognize errors when they occurred and made it much easier for them to work within their systems when things went wrong. Second, when museum employees wished to be proactive and try to solve problems that had not yet been detected, it was their familiarity with their systems that helped them do so. As time went on, they felt even more comfortable developing new ways of interacting with these databases which constantly improved their ability to recover from

errors. Although they remained unable to prevent errors from occurring, museum staff members knew that their abilities to understand and manipulate their systems would help them recover from errors before they caused irrevocable damage in the museum's collections databases.

This finding reinforces the notion that by solving errors in information systems, information professionals are doing much more than correcting mistakes: they are learning from errors and improving their abilities to work with information systems [11]. If, however, these systems are so complicated, so rigidly controlled, and so difficult to understand that their users are unable to feel comfortable working within the systems' constraints, then it is very unlikely that those users will feel confident enough to begin the process of solving problems. By making information systems more transparent, therefore, we not only encourage users to take an active role in error recovery, we improve their overall understanding of the processes and activities inherent to the information organization.

5.2 Error recovery is encouraged when collections databases promote the use of mechanisms that track changes to records over time, allowing users to backtrack errors to discover their source and potential resolution.

An important feature of the information systems used by the museum professionals to support their packing processes was that they included mechanisms to track changes to records and other information resources over time. Frequently, these came "built in," thanks to the default capabilities of paper forms (data on paper are not automatically lost when changes are made) or electronic databases (certain field or record-level changes and other statistical data can be recorded automatically in database systems). Additionally, the museum's packing policies encouraged museum staff members to carefully track changes they made to the museum's information systems. Employees in Collections were asked to make edits on existing paper forms rather than rewriting new data on new forms; employees in Registration were asked to document changes to records by adding summaries to comment fields in the collections databases. Without these data, museum staff members would have had many more problems detecting and correcting errors in their collections databases.

That museum employees had access to data documenting changes to their information systems proved valuable in every one of the 627 error transaction logs analyzed in this study. As museum staff members worked to solve packing problems, their first steps were frequently to backtrack all changes made to all records, paper and electronic, until they arrived at the source of the error. This was especially useful in the case of latent errors, where the original error might lie hidden in a box far removed from the box where the error was at last manifested. Without data tracking changes to records over time, museum staff members would have had to make many more guesses about potential sources of error, and it is likely that many packing problems would have gone unsolved.

This finding echoes results from earlier studies which illustrate the importance of collecting data about processes and other changes when building robust, sociotechnical systems [12]. Faced with errors in systems that do not include or encourage such capabilities, users often know only that a piece of data is wrong; they generally do not know why it is wrong, how it became wrong, or what they can do to solve the problem. Without that knowledge, users do not know where to

begin, and may even choose to ignore the error than deal with the frustration of tracking down a problem using insufficient data. By incorporating features into information systems that automatically track changes to records and other resources over time, we can encourage users to become more involved in solving problems as the task of tracking down the source of error becomes less onerous.

5.3 Error recovery is encouraged when collections databases provide simple and accessible communication channels that encourage users to document their attempts to detect and correct errors, clarifying complex activities in multi-user environments.

While they worked to solve errors in their collections databases, museum staff members benefited from the fact that the museum's information systems provided simple and accessible mechanisms that allowed the systems' users to communicate with each other. Just as with the mechanisms to track changes above, these "communication channels" were both an inherent part of the technologies used at the museum and an explicit part of the policies that governed the packing process at the museum. With the paper Packing Data Sheets, it was a simple matter to write a note in the margin of the form or attach a sticky note that explained something unusual about a particular artifact or particular box. Likewise, with the electronic Packing Database, comments fields were provided where museum staff members could enter questions about possible errors or suggestions for potential solutions. Museum staff members at all levels were explicitly encouraged to make use of these communication channels when packing the museum and solving packing problems. Thanks to the existence of such simple communication mechanisms, museum employees were better able to work together to solve complex problems involving multiple personnel, artifacts, and boxes.

As observed in the above examples, museum staff members depended on their abilities to communicate with each other in order to solve problems. They needed suitable mechanisms which supported collaborative problem solving over potentially difficult amounts of time and space. In solving each error, museum staff members had no way of knowing how far back into the past an error might reach, nor how many museum employees might have been involved with the error in some way or another, nor how many other problems might be related to any one error. Given the number of undergraduate employees involved in this project, their high rate of turnover, and the time lag between the creation of a latent error and its discovery, it was very likely that students involved when errors originally occurred would not be around when those errors were resolved. It was extremely important, therefore, that all attempts to detect and correct errors be carefully documented and that relevant activities be properly communicated from individual to individual, and from box to box.

This finding reinforces the argument that successful, robust information systems include features to document and track conversations about errors and other problem solving activities [3; cf. 45]. When problems are so complex that they cannot be solved by one individual at one time, it becomes extremely important that information systems provide mechanisms for communicating problems and potential solutions among the systems' users. Without that ability, users may feel overwhelmed in the face of errors, and cut off from their colleagues and co-workers. By providing users with simple and accessible communication channels which allow them to

communicate their problem-solving activities easily, information systems will encourage users to become more involved in the error recovery practices of the information organization.

6.0 Conclusion: Encouraging a Culture of Error Recovery

To keep this article at a reasonable length, we necessarily had to oversimplify certain key error management concepts: the actual relationship between error prevention and error recovery, for instance, is not as black and white as this article may have sometimes implied; certainly, lessons learned from error recovery procedures have the potential to inform and improve the error prevention process and vice versa. Along these same lines, there are many different types of error recovery, and this article frequently had to blur the distinction between the formal process of information professionals actively looking for known errors in their collections databases vs. the more informal process of serendipitous error discovery during the daily use of these same systems. Finally, it was impossible to discuss the very important issue of the roles different types of system users might play in the error recovery process, in particular the differences between the users who work in as employees in any given information organization and the users who are simply making use of publicly-available resources.

The need to examine these issues in greater detail lays the groundwork for future research in this area. There may be other factors not documented in this study that contribute to error recovery processes in information organizations. The three factors identified here are simply those that emerged from the data analysis as we explored how museum staff members dealt with their error-prone environment; it is certainly not the contention of this article that these factors are the only factors worth considering, or that attempts to implement one factor in an information system would be worthless without the other two. Follow-up studies will be necessary to explore the relative value of these factors for different types of information organizations.

Additional research will also be necessary to explore the relative difficulty of implementing these factors in collections databases in different organizations, especially in situations where different types of error management activities occur and different types of users regularly access the relevant information systems. To overcome these difficulties, it is likely that encouraging these factors in most collections databases may require the development of new technologies to facilitate the creation of transparent systems with automatic tracking of changes and accessible communication channels usable in many different scenarios. While this will undoubtedly be very challenging, an even more challenging problem may lie in the fact that the ability of these factors to encourage error recovery practices is predicated on the assumption that such activities are supported and encouraged in information organizations.

The primary reason these three factors were able to encourage error recovery procedures in the Spurlock Museum was that the culture in which the museum's employees worked encouraged such behaviors. There was a very important mindset in place at this museum; museum staff members were dedicated to the need to solve packing problems, no matter how minor the error. It will do little good, naturally, for developers to design collections databases that support error recovery procedures if they are used by employees within an organizational culture where detecting and correcting errors is not considered a valuable part of their day-to-day jobs and

perhaps even considered a waste of time. While the topic of error recovery and organizational culture deserves an entire article of its own, we feel that a few words are warranted here.

There have been arguments about the importance of correcting errors found in collections databases in libraries and other information organizations. Many have argued that serious errors tend to be so few and far between that fixing these errors is not worth the time required to do so; in the case of typographic errors, for instance, researchers have argued that few are so serious that users cannot work around these errors in their use of the system [62]. Unless error rates in collections databases are so high that they continually cause trouble for the systems' users, therefore, why worry about encouraging error recovery procedures? If, say, less than one percent of a database's records contain errors, is it really worth the time required to fix those errors, especially since we know that 100% perfection is an unachievable goal?

There is a simple answer to this question. As this article has demonstrated, information systems encouraging error recovery have a much greater potential value than merely helping employees to correct the odd, unexpected error that sneaks through already existing error prevention mechanisms. By supporting and encouraging error recovery in collections databases, employers are actually supporting and encouraging a culture where employees are rewarded for taking an active role in improving the quality of data across the entire organization. The attitude that there is no reason to bother correcting a small number of errors can breed an environment where employees are more likely to look the other way when errors occur, even if they were in a position to correct those errors and regardless of the severity of the error or their desire to fix it. In contrast, the attitude that one should not let an opportunity to correct mistakes slide by, even knowing full well that correcting all mistakes is impossible, can encourage employees to improve data quality in collections databases and to believe that developing more robust information systems is a worthwhile goal to pursue.

The implications of this article, therefore, are generalizable for all information organizations at multiple levels of implementation. First, they demonstrate how error recovery procedures, if properly implemented and encouraged in collections databases, can help the users of these systems play a role in detecting and correcting unexpected errors, solving a variety of problems for libraries, museums, and archives. Second, they document how collections databases which are flexible, open, and designed to work with errors tend to be more robust and more likely to encourage error recovery than systems which are rigid, closed, and designed to prevent errors. Finally, they illustrate how information organizations concerned with finding the best error management techniques for their collections databases will reap rewards that go beyond a simple ability to correct errors.

Acknowledgements

The author would like to acknowledge the many contributions of the staff and students of the Spurlock Museum. Without their hard work, this project would never have been possible.

References

1. Medawar, Katia. "Database Quality: A Literature Review of the Past and a Plan for the Future." *Program* 29, no. 3 (1995): 257-272.
2. O'Neill, Edward T.; and Vizine-Goetz, Diane. (1988). "Quality Control in Online Databases." *Annual Review of Information Science and Technology* 23 (1988): 125-156.
3. Robbin, Alice; and Frost-Kumpf, Lee. "Extending Theory for User-Centered Information Services: Diagnosing and Learning from Error in Complex Statistical Data. *Journal of the American Society for Information Science* 48, no. 2 (1997): 96-121.
4. Adams, Stephen. "Information Quality, Liability, and Corrections." *Online* 27, no. 5 (2003): 16-22.
5. Beall, Jeffrey; and Kafadar, Karen. "The Effectiveness of Copy Cataloguing at Eliminating Typographical Errors in Shared Bibliographic Records." *Library Resources & Technical Services* 48, no. 2 (2004): 92-101.
6. Klein, Barbara D. "The Detection of Data Errors in Computer Information Systems: Field Interviews with Municipal Bond Analysts." *Information Resources Management Journal* 13 no. 3, (2000): 23-32.
7. Bade, David. "The Creation and Persistence of Misinformation in Shared Library Catalogs: Language and Subject Knowledge in a Technological Era." Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, 2002.
8. Wang, Richard Y.; and Strong, Diane M. "Beyond Accuracy: What Data Quality Means to Data Consumers." *Journal of Management Information Systems* 12, no. 4 (1996): 5-34.
9. Fox, Christopher; Levitin, Anany; and Redman, Thomas. "The Notion of Data and its Quality Dimensions." *Information Processing & Management* 30, no. 1 (1994): 9-19.
10. Redman, Thomas C. *Data Quality: Management and Technology*. New York: Bantam Books, Inc., 1992.
11. Seifert, Colleen M.; and Hutchins, Edwin L. "Error as Opportunity: Learning in a Cooperative Task." *Human-Computer Interaction* 7, no. 4 (1992): 409-435.
12. Twidale, Michael B.; and Marty, Paul F. "Coping with Errors: The Importance of Process Data in Robust Sociotechnical Systems." In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work* (pp 269-278). Chapel Hill: ACM Press, 2000.
13. Armstrong, Chris J. "Database Information Quality." *Library & Information Briefings* 62 (1995): 1-14.
14. Ballard, Terry. "Spelling and Typographical Errors in Library Databases." *Computers in Libraries* 12, no. 6 (1992): 14-17.
15. Marty, Paul F. "Museum Informatics and the Evolution of an Information Infrastructure in a University Museum." Ph.D. diss., University of Illinois at Urbana-Champaign, 2002.
16. Laudon, Kenneth C. "Data Quality and Due Process in Large Interorganizational Record Systems." *Communications of the ACM* 29 (1986): 4-11.
17. Morey, Richard C. "Estimating and Improving the Quality of Information in a MIS." *Communications of the ACM* 25 (1982): 337-342.
18. Jacsó, Peter. "Searching for Skeletons in the Database Cupboard Part I: Errors of Omission." *Database* 16, no. 1 (1993): 38-49.
19. Jacsó, Peter. "April Fool's Day Follies." *Information Today* 19, no. 4 (2002): 34-35.
20. Redman, Thomas C. "The Impact of Poor Data Quality on the Typical Enterprise." *Communications of the ACM* 41, no. 2 (1998): 79-82.

21. Nielsen, Ralph; and Pyle, Jan M. "Lost Articles: Filing Problems with Initial Articles in Databases." *Library Resources & Technical Services* 39, no. 3 (1995): 291.
22. Redman, Thomas C. *Data Quality for the Information Age*. Norwood, MA: Artech House, 1996.
23. Wang, Richard Y. "A Product Perspective on Total Data Quality Management." *Communications of the ACM* 41, no. 2 (1998): 58-65.
24. Klein, Barbara D. "How Do Actuaries Use Data Containing Errors? Models of Error Detection and Error Correction." *Information Resources Management Journal* 10, no. 4 (1997): 27-36.
25. Wand, Yair; and Wang, Richard Y. "Anchoring Data Quality Dimensions in Ontological Foundations." *Communications of the ACM* 39, no. 11 (1996): 86-95.
26. Wang, Richard Y.; Reddy, Martin P.; and Kon, Henry B. "Toward Quality Data: An Attribute-Based Approach." *Decision Support Systems* 13, no. 3/4 (1995): 349-372.
27. Ballou, Donald P.; and Tayi, Giri K. "Enhancing Data Quality in Data Warehouse Environments." *Communications of the ACM* 42, no. 1 (1999): 73-78.
28. Norman, Donald A. "Design Rules Based on Analyses of Human Error." *Communications of the ACM* 4 (1983): 254-258.
29. Ackerman, Marc S. "The Intellectual Challenge of CSCW: The Gap Between Social Requirements and Technical Feasibility." In *Human Computer Interaction in the New Millennium*, edited by John M. Carroll (pp. 303-324). New York: ACM Press, 2000.
30. Strong, Diane M. & Miller, Steven M. (1995). "Exceptions and Exception Handling in Computerized Information Processes." *ACM Transactions on Information Systems* 13, no. 2 (1995): 206-233.
31. Abbott, Kenneth R.; and Sarin, Sunil K. "Experiences with Workflow Management: Issues for the Next Generation." In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work* (pp. 113-120). Chapel Hill: ACM Press, 1994.
32. Gasser, Les. "The Integration of Computing and Routine Work." *ACM Transactions on Office Information Systems* 4, no. 3 (1986): 205-225.
33. Gerson Elihu M.; and Star, Susan L. "Analyzing Due Process in the Workplace." *ACM Transactions on Office Information Systems* 4, no. 3 (1986): 257-270.
34. Grinter, Rebecca E. "Supporting Articulation Work Using Software Configuration Management Systems." *Computer Supported Cooperative Work* 5 (1996): 447-465.
35. Schmidt, Kjeld; and Bannon, Liam. "Taking CSCW Seriously: Supporting Articulation Work." *Computer Supported Cooperative Work* 1, no. 1 (1992): 7-40.
36. Star, Susan L.; and Strauss, Anselm. "Layers of Science, Arenas of Voice: The Ecology of Visible and Invisible Work." *Computer Supported Cooperative Work* 8 (1999): 9-30.
37. Suchman, Lucy A. "Supporting Articulation Work." In *Computerization and Controversy*, edited by Rob Kling (pp. 407-423). San Diego: Academic Press, 1996.
38. Winograd, Terry; and Flores, Fernando. *Understanding Computers and Cognition: A New Foundation for Design*. Reading, MA: Addison-Wesley Publishing Company, 1987.
39. Robinson, Mike. "Design for Unanticipated Use..." In *Proceedings of the European Conference on Computer-Supported Cooperative Work* (pp. 187-202). Dordrecht: Kluwer Academic Publishers, 1993.
40. Dourish, Paul. "The Appropriation of Interactive Technologies: Some Lessons from Placeless Documents." *Computer Supported Cooperative Work*, 12, no. 4 (2003): 465-490.

41. Pekkola, Samuli. "Designed for Unanticipated Use: Common Artifacts as Design Principle for CSCW Applications." In *Proceedings of the ACM Conference on Supporting Group Work* (pp. 359-368). New York: ACM Press, 2003.
42. Rogers, Yvonne. "Exploring obstacles: Integrating CSCW in Evolving Organizations." In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work* (pp. 67-66). Chapel Hill: ACM Press, 1994.
43. Dourish, Paul; Holmes, Jim; MacLean, Allen; Marquardsen, Pernille; and Zbyslaw, Alex. "Freeflow: Mediating Between Representation and Action in Workflow Systems." In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work* (pp. 190-198). Chapel Hill: ACM Press, 1996.
44. Luff, Paul; Hindmarsh, Jon; and Heath, Christian. *Workplace Studies: Recovering Work Practice and Informing System Design*. Cambridge: Cambridge University Press, 2000.
45. Auramäki, Esa; Robinson, Mike; Aaltonen, Ann; Kovalainen, Mikko; Liinamaa, Arja; Tuuna-Vaiska, Taina. "Paperwork at 78kph." In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work* (pp. 370-379). Chapel Hill: ACM Press, 1996.
46. Blumenthal, Richard; and Nutt, Gary J. "Supporting Unstructured Workflow Activities in the Bramble System." In *Proceedings of the Conference on Organizational Computing Systems* (pp. 130-138). New York: ACM Press, 1995.
47. Saastamoinen, Heikki; and White, George M. "On Handling Exceptions." In *Proceedings of the Conference on Organizational Computing Systems* (pp. 302-310). New York: ACM Press, 1995.
48. Trigg, Randall H.; and Bodker, Susanne. "From Implementation to Design: Tailoring and the Emergence of Systematization in CSCW." In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work* (pp. 45-54). Chapel Hill: ACM Press, 1994.
49. Dourish, Paul. "Developing a Reflective Model of Collaborative Systems." *ACM Transactions on Computer-Human Interaction* 2, no. 1 (1995): 40-63.
50. Bowker, Geoffrey; Star, Susan L.; Turner, William, & Gasser, Les. *Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide*. Mahwah, NJ: Lawrence Erlbaum Associates, 1997.
51. Davis, Gordon B.; Neter, John; and Palmer, Richard. "An Experimental Study of Audit Confirmation." *Journal of Accountancy* 123, no. 6 (1967): 36-44.
52. Ricketts, John A. "Powers-of-Ten Information Biases." *MIS Quarterly* 14 (1990): 63-77.
53. Kovalainen, Mikko; Robinson, Mike; and Auramäki, Esa. "Diaries at Work." In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work* (pp. 49-58). New York: ACM Press, 1998.
54. Hill, Will; and Hollan, Jim. "History-Enriched Digital Objects: Prototypes and Policy Issues." *The Information Society*, 10, no. 2 (1994): 139-145.
55. Mariani, John A.; and Rodden, Tom. "The Impact of CSCW on Database Technology." In *the International Workshop on CSCW* (pp. 146-156). edited by K. Gorling and C. Sattler. Berlin: Informatik Informationen-Reporte, 1991.
56. Pycock, James; Palfreyman, Kevin; Allanson, Jen; and Button, Graham. "Representing Fieldwork and Articulating Requirements Through VR." In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work* (pp. 383-392). New York: ACM Press, 1998.
57. Marty, Paul F.; and Twidale, Michael B. "Unexpected Help with your Web-based Collections: Encouraging Data Quality Feedback from your Online Visitors. In *Proceedings*

- of Museums and the Web*, edited by David Bearman and Jennifer Trant. Pittsburgh, PA: Archives and Museum Informatics, 2000.
58. Beall, Jeffrey. "The Dirty Database Test." *American Libraries* 22 (1991): 97.
 59. Dwyer, Jim. "The Catalogers' 'Invisible College' at Work: The Case of the Dirty Database Test." *Cataloging and Classification Quarterly* 14, no. 1 (1991): 75-82.
 60. Bourne, Charles P. "Frequency and Impact of Spelling Errors in Bibliographic Databases." *Information Processing and Management* 13, no 1. (1977): 1-12.
 61. Gardner, Sylvia A. "Spelling Errors in Online Databases: What the Technical Communicator Should Know." *Technical Communication* 39 (1992): 50-53.
 62. Randall, Barbara N. "Spelling Errors in the Database: Shadow or Substance?" *Library Resources & Technical Services* 43, no 3. (1999): 161-169.
 63. Mintz, Anne P. "Quality Control and the Zen of Database Production." *Online* 14, no. 6 (1990): 15-23.
 64. Orr, Ken. "Data Quality and Systems Theory." *Communications of the ACM* 41, no. 2 (1998): 66-71.
 65. Davis, Carol. "Results of a Survey on Record Quality in the OCLC Database." *Technical Services Quarterly* 7, no. 2 (1989): 43-53.
 66. Ojala, Marydee. "The Finns Again Wake: The Wake Up Call to Information Quality." *Information Today* 10, no. 3 (1993): 41-42.
 67. Marty, Paul F. "Museum Informatics and Information Infrastructures: Supporting Collaboration Across Intra-Museum Boundaries." *Archives and Museum Informatics* 13, no. 2 (1999): 169-179.
 68. Marty, Paul F. "Museum Informatics: Sociotechnical Infrastructures in Museums." *Bulletin of the American Society for Information Science* 26, no. 3 (2000): 22-24.
 69. Stake, Robert E. *The Art of Case Study Research*. Thousand Oaks, CA: Sage Publications, 1995.
 70. Yin, Robert K. *Case Study Research, Design and Methods*. Newbury Park: Sage Publications, 2002.
 71. Benson, Douglas; and Hughes, John A. *The Perspective of Ethnomethodology*. London: Longman, 1983.
 72. Hutchins, Edwin. *Cognition in the Wild*. Cambridge: MIT Press, 1995.
 73. Strauss, Anselm; and Corbin, Juliet. *Basics of Qualitative Research*. Thousand Oaks, CA: Sage Publications, 1998.
 74. Campbell, John P. "Modeling the performance prediction problem in industrial and organizational psychology." In *Handbook of Industrial and Organizational Psychology* (pp. 687-732), edited by M.D. Dunnette and L.M. Hough. Palo Alto, CA: Consulting Psychologists Press, Inc., 1990.
 75. Dervin, Brenda. "Information as a User Construct: The Relevance of Perceived Information Needs to Synthesis and Interpretation." In *Knowledge Structure and the Implications for Synthesis and Interpretation* (pp. 155-183), edited by S. A. Ward and L. J. Reed. Philadelphia: Temple University Press, 1983.
 76. Morris, Ruth C. "Toward a User-Centered Information Service." *Journal of the American Society for Information Science* 45, no. 1 (1994): 20-30.
 77. Wilson, Patrick. "Unused Relevant Information in Research and Development." *Journal of the American Society for Information Science* 46, no. 1 (1995): 45-51.

78. Webber, Douglas J. "The Distribution and Use of Policy Knowledge in the Policy Process." *Knowledge and Policy: The International Journal of Knowledge Transfer and Utilization* 4, no. 4 (1991): 6-35.
79. Marty, Paul F. "Factors Influencing the Co-Evolution of Computer-Mediated Collaborative Practices and Systems: A Museum Case Study." *Journal of Computer-Mediated Communication* 10, no. 4 (2005): in press.

Packing Data

Pack Type B	Box Number: 1340	Packed By: TEO	Date: JAN 25, 2000
Mat Type M	Box type: <input type="checkbox"/> Constructed <input checked="" type="checkbox"/> Pre-fabricated	Supervised By: Chub	Date: 1/25/2000
Display AMN	Packing Type: <input checked="" type="checkbox"/> Soft Pack: Bolster <input type="checkbox"/> Cavity Pack <input checked="" type="checkbox"/> Soft Pack: Tray <input type="checkbox"/> Crate: Slat <input checked="" type="checkbox"/> Soft Pack: Wrap <input type="checkbox"/> Crate: Full	IPM: <input type="checkbox"/> Deinfested, Active <input type="checkbox"/> Deinfested, Residual <input type="checkbox"/> Potential, High <input checked="" type="checkbox"/> Potential, Low <input type="checkbox"/> None	
Problems	Material Types: <input checked="" type="checkbox"/> Organic <input type="checkbox"/> Textiles <input checked="" type="checkbox"/> Inorganic <input checked="" type="checkbox"/> Paper/Book		

Box Location

Room Number: 484	Zone: 22	Cabinet:	Shelf:
Moved into Storage Zone by: SAS	Date: 5/23/00		
Loc. Entered into Pkng Dbse by: BN	Date: 5/24/00		

Packing Remarks

Label Check ✓ OK-MB 5/24/00	
---	--

Box Contents (Accession Numbers)

Total number of items 18/33

Acc.# not found in database

1921.01.0007	1926.02.0205
1968.05.0027	1922.01.0031
FOLDER 1 (15 items)	1911.02.0018
1976.24.0226B	1900.53.0019
1926.02.0095	1926.02.0191
1921.01.0003	
1927.01.0021	
1926.02.0198	
1900.11.0080	
1954.04.0013	✓ OK SAS 2/16/00
1922.01.0025	Duplicate
1922.10.0011	✓ OK SAS 2/16/2000
1954.04.0013	omit duplicate entry ✓ SAS 2/16/00
1922.10.0019	

Database has this in Box 0531

17/32
✓ SAS 2/16/2000

Entered by: fm (bpd) MB Checked by: _____
 Date: 1-28-2000 Date: _____

Figure 1: Packing Data Sheet for Box 1340

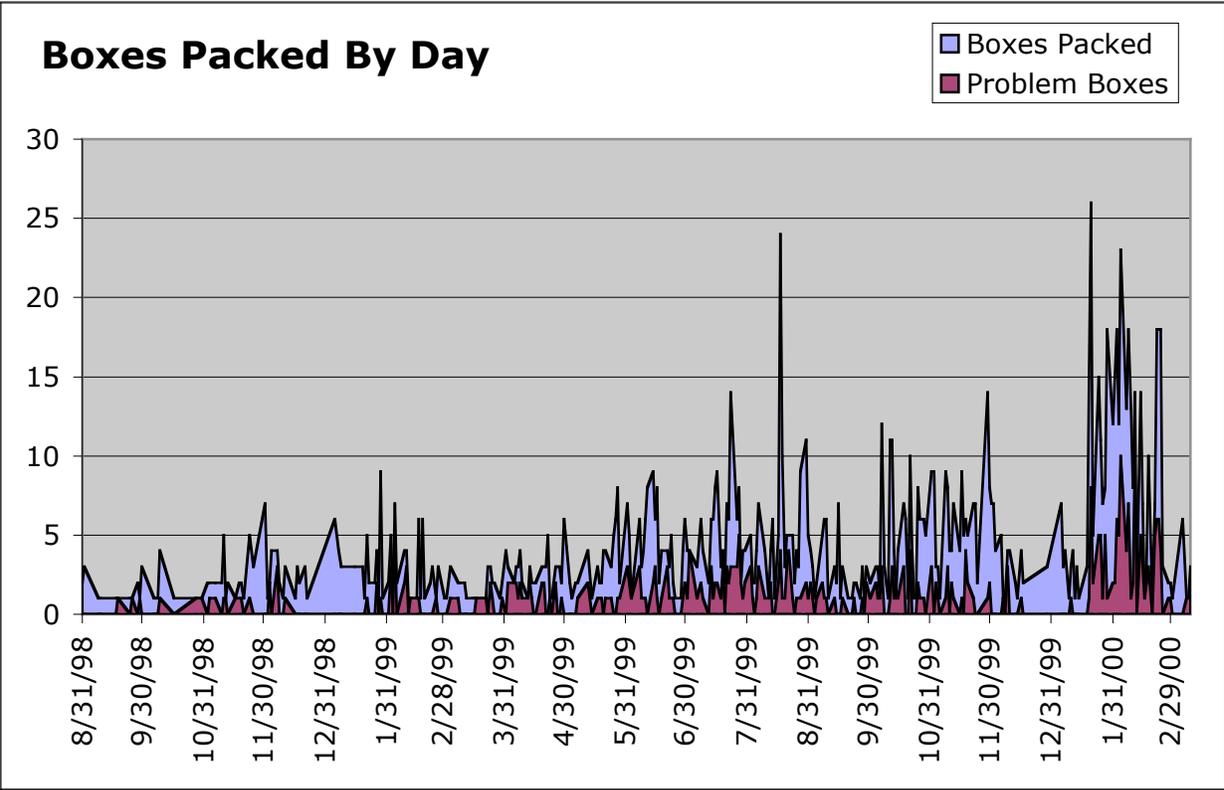


Chart 1: Problem Boxes by Boxes Packed per Day

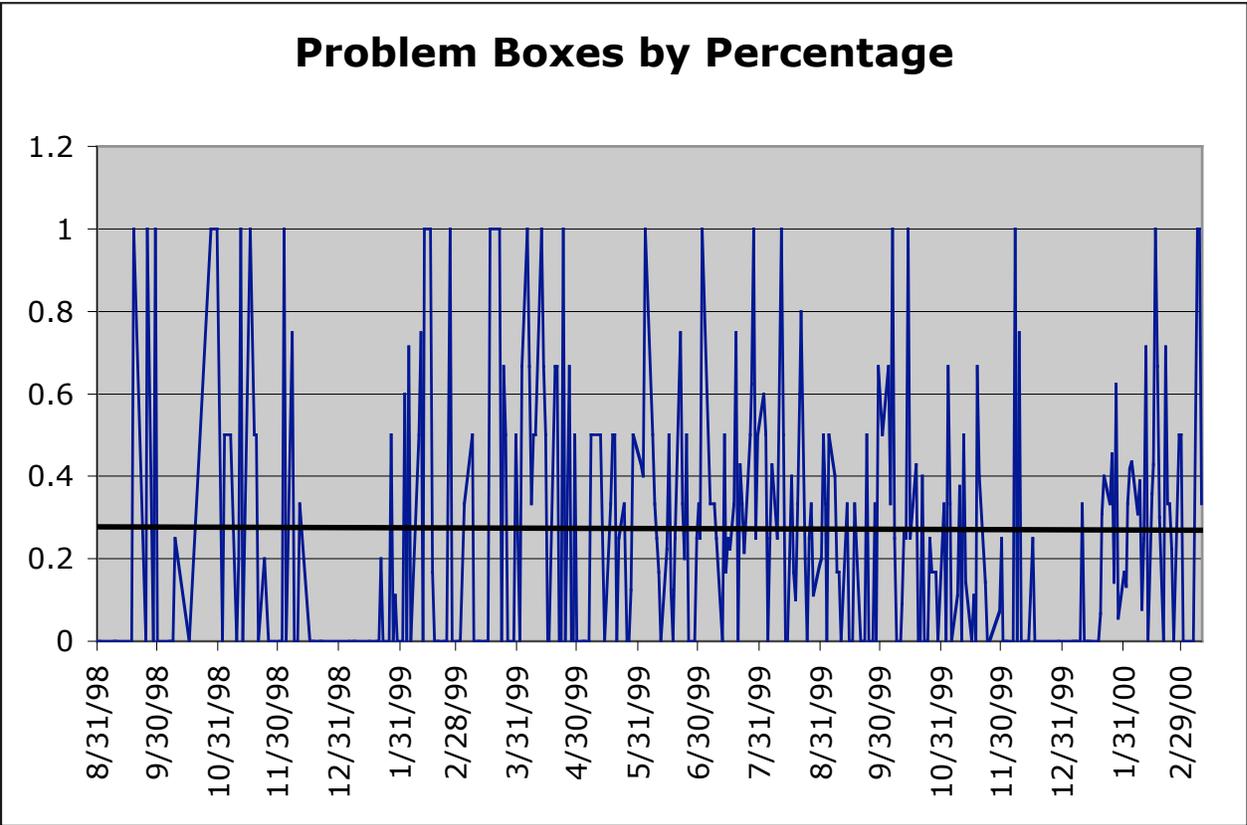


Chart 2: Problem Boxes by Percentage of Boxes Packed each Day

	Artifacts	Boxes
<i>Total Records Evaluated</i>	27,166	1408
<i>Records with Errors</i>	627	373
<i>Error Rates</i>	2.3%	26.5%

Table 1: Errors and Error Rates by Artifacts and Boxes Packed

Errors Detected During Data Entry	489
<i>No Such Number</i>	210
<i>Already Packed</i>	221
<i>Duplicate Number</i>	58
Errors Recovered After Data Entry	138
<i>Latent Errors</i>	138
Total Error Transaction Logs Evaluated	627

Table 2: Error Transaction Logs by Type

Errors Detected & Corrected During Packing	627
Errors Missed & Not Corrected During Packing	4
Error Recovery Rate	99.37%

Table 3: Error Recovery Rates