IDEA GROUP PUBLISHING
701 E. Chocolate Avenue, Hershey PA 17033-1117, USA ITB7398
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

Chapter XVI

Some Evidence on the Detection of Data Errors

Barbara D. Klein
University of Michigan, Dearborn, USA

Data stored in organizational databases have a significant error rate. As computerized databases continue to proliferate, the number of errors in stored data and the organizational impact of these errors are likely to increase. The impact of data errors on business processes and decision making can be lessened if users of information systems are able and willing to detect and correct data errors. However, some published research suggests that users of information systems do not detect data errors. This paper reports the results of a study showing that municipal bond analysts detect data errors. The results provide insight into the conditions under which users in organizational settings detect data errors. Guidelines for improving error detection are also discussed.

INTRODUCTION

Data stored in organizational databases have an error rate between 1 and 10% (Laudon, 1986; Madnick & Wang, 1992; Morey, 1982; Redman, 1992). As computerized databases continue to proliferate and as organizations become increasingly dependent upon these databases to support business processes and decision making, the number of errors in stored data and the organizational impact of these errors are likely to increase. For example, strategies such as total quality management may be difficult to implement if the required data are not of adequate quality (Fox, Levitin & Redman, 1993; Madnick & Wang, 1992; Redman, 1995).

Possible approaches for managing data errors in organizations include: (1) validating data during input or storage (e.g., Morey, 1982) and (2) relying on detection and correction of errors by end users. While useful, automated approaches to data validation do not generally yield completely accurate data. Indeed, Orr (1998)

This chapter appears in the book, Advanced Topics in Information Resources Management by Mehdi Khosrow-Pour.

argues that unused data will become inaccurate over time. The best approach to reducing data errors in most organizations will include both automated data validation and user detection of errors. This paper examines user detection of data errors in the profession of municipal bond analysis.

The remaining sections of this paper present (1) a review of prior research, (2) a theory of error detection, (3) the research design, (4) the empirical results, and (5) a discussion of the implications of the results.

BACKGROUND

Several conclusions can be drawn from the literature on data quality. First, while no definition of data quality has been completely accepted, there is agreement that accuracy, currency, and completeness are important aspects of data quality (Agmon & Ahituv, 1987; Davis & Olson, 1985; Fox et al., 1993; Huh, Keller, Redman & Watkins, 1990; Madnick & Wang, 1992; Wand & Wang, 1996; Zmud, 1978). A broader perspective on data quality, which is gaining widespread acceptance among data quality researchers and practitioners, includes 15 dimensions of data quality: believability, accuracy, objectivity, reputation, value-added, relevancy, timeliness, completeness, appropriate amount of data, interpretability, ease of understanding, representational consistency, concise representation, accessibility, and access security (Strong, Lee & Wang, 1997; Wang & Strong, 1996). Second, error rates significantly greater than zero have been found in several studies (Ham, Losell & Smieliauskas, 1985; Johnson, Leitch & Neter, 1981; Knight, 1992; Laudon, 1986; Stone & Bublitz, 1984). Third, researchers disagree about the extent to which efforts to rid databases of errors should be undertaken. Some advocate methods designed to completely eliminate errors from databases (Janson, 1988; Naus, 1975; Parsaye & Chignell, 1993; Svanks, 1988; Wang, Lee, Pipino & Strong, 1998). Others propose methods for allocating limited resources to data quality initiatives (Ballou & Pazer, 1987; Ballou & Tayi, 1989; Ballou, Pazer, Belardo & Klein, 1987; Bowen, 1992; Paradice & Fuerst, 1991). Fourth, approaches for using imperfect data have been proposed (Ballou & Pazer, 1985, 1987, 1995; Bansal, Gaba & Winkler, 1992; Garfinkel, Kunnathur & Liepins, 1986; Kauffman & Weitz, 1993; O'Leary, 1993; O'Neill & Vizine-Goetz, 1988).

Related research has examined errors in spreadsheets (e.g., Panko, 1998) and user assessments of the quality of Internet-based information (Alexander & Tate, 1999; Klein, 2001; Rieh & Belkin, 1998).

The early literature on data quality suggests that users are not effective at detecting errors in data. Davis, Neter and Palmer (1967) found that half the people asked to verify their banking account information with imbedded errors failed to detect the errors. Laudon (1986) found that users of criminal information systems rarely detect errors in this data. Ricketts (1990) found that over 90% of the people participating in a laboratory experiment failed to detect a data error in production planning reports. Much of the literature on data quality assumes that humans will fail to detect data errors and argues that resources should be allocated to the improvement of data quality as data are input to databases (e.g., Redman, 1992, 1995).