

Attention to data quality is a critical issue in all areas of information resources management. A recent article in the Wall Street Journal (7/13/98) relates the domino effect that occurs when erroneous information is typed into a central database. A new airport in Hong Kong suffered catastrophic problems in baggage handling, flight information, and cargo transfer. The ramifications of the dirty data were felt throughout the airport. Flights took off without luggage, airport officials tracked flights with plastic pieces on magnetic boards, and airlines called confused ground staff on cellular phones to let them know where even more confused passengers could find their planes (Arnold, 1998). The new airport had been depending on the central database to be accurate. When it wasn't, the airport paid the price in terms of customer satisfaction and trust.

Data warehousing is emerging as the cornerstone of an organization's information infrastructure. It is imperative that the issue of data quality be addressed if the data warehouse is to prove beneficial to an organization. Corporations, government agencies and not-for-profit groups are all inundated with enormous amounts of data. The desire to use this data as a resource for the organization has increased the move towards data warehouses. This information has the potential to be used by an organization to generate greater understanding of their customers, processes, and the organization itself.

There potential to increase the usefulness of data by combining it with other data sources is great. But, if the underlying data is not accurate, any relationships found in the data warehouse will be misleading. For example, most payroll systems require a social security number when setting up an employee file. If no number is available when the file is set up, an incorrect number may be used, such as 111-11-1111, in order to facilitate payroll processing. The intention is that the numbers would be changed when the correct social security number is obtained. If the numbers are not changed, then some relationship may exist in the database, but the relationship would be misleading because the underlying data is inaccurate.