Evaluating the Quality and Usefulness of Data Breach Information Systems

Benjamin Ngugi, Suffolk University, USA

Jafar Mana, Suffolk University, USA

Lydia Segal, Suffolk University, USA

ABSTRACT

As the nation confronts a growing tide of security breaches, the importance of having quality data breach information systems becomes paramount. Yet too little attention is paid to evaluating these systems. This article draws on data quality scholarship to develop a yardstick that assesses the quality of data breach notification systems in the U.S. at both the state and national levels from the perspective of key stakeholders, who include law enforcement agencies, consumers, shareholders, investors, researchers, and businesses that sell security products. Findings reveal major shortcomings that reduce the value of data breach information to these stakeholders. The study concludes with detailed recommendations for reform.

Keywords: Data Breach, Data Quality, Data Security, Notification Laws, Usefulness

INTRODUCTION

Data breaches can jeopardize the livelihoods of hundreds of thousands of people (Javelin Strategy and Research, 2010a). Each breach can expose large numbers of customer records. A singular computer intrusion at TJX Companies Inc. in 2007, for instance, resulted in the loss of over 45 million customer records (Privacy Commissioner -Canada & Information & Privacy Commissioner-Alberta, 2007). If a fraction of these records had fallen into the wrong hands, a major identity theft crisis would have been under way. Indeed, security breaches are a leading cause of identity theft, the number one consumer complaint from 2000 to 2010, consisting of 19% of the overall complaints (Federal Trade Commission, 2010, 2011). Data breaches, moreover, are increasing. Between 2005 and 2009, the number of data breaches in the U.S. rose from 157 to 498 (Identity Theft Resource Center, 2010c). Total annual identity fraud has been rising from $45 in 2007 to $54 billion in 2009 (Javelin Strategy and Research, 2010b). Despite this, however, there is no common yardstick by which to evaluate data breach systems.

DOI: 10.4018/jisp.2011100103

The goal of this study is to develop a yardstick and evaluate how well current data breach notification systems are meeting stakeholder needs.

The remainder of this article is organized as follows: Part II reviews the relevant literature. Part III sets forth the methodology. Part IV creates a yardstick by which to evaluate data breach notification systems. Part V applies that yardstick to evaluate data breach information systems. Part VI offers recommendations for reform, while Part VII summarizes the contributions and offers suggestions for future research.

LITERATURE REVIEW

The data quality literature has long discussed the importance of quality (Juran & Godfrey, 1999; Wand & Wang, 1996; Wang, Storey, & Firth, 1995). Decisions made on the basis of corrupt or inferior data will be skewed, with potentially costly consequences (Baltzan & Phillips, 2009; Fisher, Chengalur-Smith, & Ballou, 2003). As Baltzan and Phillips (2009) observe, "decisions are only as good as the quality of data breach information used to make the decisions."

Researchers have devoted much energy to investigating how to evaluate information for quality. One of the most prominent such scholars, professor and Director of the MIT Information Quality Program Richard Wang, has written several seminal papers on the subject. In one such paper, Wang and Strong (1996) develop a conceptual framework designed to capture "the aspects of data quality that are important to consumers" (Wang & Strong, 1996, p. 5). The framework conceives of data quality as comprising four dimensions. One dimension refers to the intrinsic factors of the data itself. Examples are the data's accuracy, objectivity, believability, and reputation – all of which go to the data's quality in their own right. The second dimension refers to contextual factors. Data quality "must be considered within the context of the task at hand" (Wang & Strong, 1996, p. 6). Contextual factors include value-added, relevance, timeliness, completeness, and appropriate amount of data. Third, the data's representational dimension includes aspects related to its format (e.g., whether it offers a concise and consistent representation) and meaning (e.g., its interpretability and the ease with which it can be understood). The last dimension is its accessibility. Data needs to be secure, while being accessible. This four-dimensional model is widely accepted by other scholars in the data quality field (Bovee, Srivastava, & Mak, 2003; Strong, Lee, & Wang, 1997).

Underlying the four dimensions of Wang and Strong's (1996) model is the concept of usefulness, a notion captured by the "fit to use" principle in the data quality literature. The importance of usefulness stems from the realization that the quality of data "depends on the actual use of data and what may be considered good data in one case (for a specific application or user) may not be sufficient in another case" (Wand & Wang, 1996). The idea is that it is critical to adopt the user's perspective on the data's fitness for whatever use the user requires the data for (Juran & Godfrey, 1999).

The concept of usefulness is built into the very definition of data quality in Wang and Strong's (1996, p. 6) model: they define data quality as data that is fit for use by data consumers. The idea of usefulness has been adopted in a number of fields, such as health surveillance (Buehler, Hopkins, Overhage, Sosin, & Tong, 2004), as part of an effort to assess data quality. Scholars have not, however, applied it yet to data breach notification systems.

Applying a generic model such as Wang and Strong's to the data breach notification area requires the use of metrics appropriate to that area. A metric is a "verifiable measure, stated in quantitative or qualitative terms and defined with respect to a reference point" (Melnyk, Stewart, & Swink, 2004 ; Payne, 2006). Metrics provide a scientific yardstick by which other similar systems can be objectively evaluated and enable a system or data set to be assessed in a way that can be defined, standardized, and systematically processed (Palmer, 2002). As Melnyk et al. (2004) and Payne (2006) note, metrics help provide control by enabling manag-