Words Are Louder Than Actions – Design Requirements for Facilitating Restorative Justice in Data Breaches

Till Ole Diesterhöft[1[0000-0002-4141-3261]], Marvin Braun[1[0000-0003-0885-3561]], Aycan Aslan[1[0000-0002-5755-9310]], Maike Greve[1[0000-0003-3766-8285]], and Lutz M. Kolbe[1[0000-0003-4852-0040]]

[1]Georg-August-Universität Göttingen, Göttingen, Germany

Abstract. Customer data breaches are distinct security incidents as data breach notification laws establish a novel communication interface to customers. Company-internal faults are revealed (e.g., neglect of security standards), customers lose trust in the company, and customer churn occurs. In response, literature developed response strategies to mitigate these losses. However, we notice that no efforts are made to provide personalized information, despite being highly promising to address further moderation of the induced damages. Drawing on restorative justice, in this ongoing research study initial meta-requirements for the development of conversational agents in data breaches are derived. We form a basis for the development of a communicative, two-way exchange artefact between customers and companies in a data breach response.

Keywords: Data Breach Response, Conversational Agents, Restorative Justice.

1 Motivation and Problem Statement

Data breaches are becoming much more of a persistent risk for companies than a rare threat. As the number of data breaches continues to rise, events that expose customer information to the public emerge as an integral part of company's everyday security response management. In 2020, Nintendo, a Japanese game company, fell victim to one of the largest cyberattacks in the gaming industry. Nintendo informed its customers via news on their website. In total, personal information of over 300,000 customers was affected [1]. Data breaches like Nintendo's possess risks for companies because they directly impact a company's stock value [2] and financial performance [3]. Additionally, by publicly disclosing a data breach, indirect costs arise. These can be attributed to decreased customer trust [4] and negative word-of-mouth [5].

Recent research has revealed that external communication after a data breach can mitigate these indirect costs [6]. Literature found that compensating customers and offering apologies can yield beneficial impacts [4, 6]. Other strategic approaches demonstrate the multidimensional nature of this emerging management challenge [3, 5]. Particularly, it could be shown that the data breach notification of each company must be tailored to the scenario and to the characteristics of the breach [2, 7]. Hence, data breach notifications constitute a strategic element of incident response manage-

ment. Companies must develop appropriate strategies and meet customer expectations in communicating data breaches [6]. Only then a mitigation of damages is possible.

However, notifications of data breaches are exclusively conveyed in a text-based form, e.g., letter or press statements [2]. This notion is prevalent both in theory [7] and practice [8]. Customers are informed about what has happened, who is affected and what actions will or have been taken by the company. Accordingly, researchers and practitioners leverage a one-way, asynchronous communication form for data breaches [9, 10]. We argue that this form of messaging contradicts the idea of tailored strategic notifications [6]. Individualized informational aspects of customers fail to be addressed by overarching, general notifications. Accordingly, individual customer needs may not be adequately met [4]. Adverse effects arise [10], which may be avoided by providing information in a selective and customer-centered manner.

Given the increasing importance of data breach communication and its hitherto neglected individual informational facets, we draw on restorative justice. The idea of restorative offers a cornerstone for establishing an innovative customer-centric way of communicating [11]. Originating as a form of justice in ancient civilizations [12], restorative justice approaches the process of reconciliation between all parties after an actor has been harmed [13]. Restorative justice aims to restore trust and deteriorated relationships by providing individualized information [13]. Central to this is a dialogue between the actors. This dialogue is supposed to establish a communication channel in which victims (e.g., customers) engage in a direct exchange with the responsible parties (e.g., companies) [13, 14]. Hence, providing an overall concept to leverage individualized information provision in data breach communication.

We reason that the concept of conversational agents (CA) fits to instantiate restorative justice in a data breach context. As dialogue-based systems, CAs are suitable for providing informative aspects in a dynamic environment [15]. Thus, given their various application possibilities and context-specific customization potentials [15, 16], CAs can be a valuable tool in the event of a data breach. Leveraging CAs is especially relevant considering the number of customers affected on average in a data breach [2]. A direct, person-to-person exchange would be virtually impossible to implement from a resource perspective. Companies could offer not only a text-based notification, but also a CA including data breach details sought by the user. Available data and information can be provided more personalized. Thus, implementing CAs in data breach communication serves to achieve two purposes: (i) Leveraging restorative justice to deliver various informative aspects, and (ii) providing a personalized communication approach that addresses the drawbacks of text-based analog notifications.

CAs are primarily used in the service sector, where various transactions are carried out. In this context, CAs assist in the transaction process and aim to conclude transactions. However, the data breach response process follows the goal of providing the customer with suitable and personalized information. It is not mandatory that a customer always achieves the same goal (i.e., transaction completion), but rather receives individual, data breach-specific experiences. Given the favorable characteristics of utilizing CAs for data breach communication and the void of alternative, comparable CA implementations, we pose the following research question: ***How should a CA be designed to be applicable in a data breach response process?***

Following the Design Science Research (DSR) paradigm [17] in an iterative manner, we aim to guide practitioners towards the use of CAs in data breach responses (in the following referred to as conversational data breach agents (CDBA)). We first identify and examine 324 actual data breach notifications. In addition, we leverage literature on meta-requirements (MR) to derive a basis for a general conceptualization of CDBA's [18, 19]. These MR provide the foundation for our next development iterations, including the derivation of design principles. In subsequent research we then aim to implement a text-based approach of CDBA in translating the MR into a technological prototype [15]. This is justified by the inherent link to textual information in data breach notifications. The interpersonal aspect of data breach response communications has been identified as a key driver of customer perceptions [5]. Hence, we situate the chatbot to be developed in the area of relationship-oriented chatbots [16]. Furthermore, we will emphasize the link to internal systems of companies as a central design feature. These contain information about affected customers of a data breach. Thus, an individualized response to data breaches can be ensured.

Our research contributes to the data breach, restorative justice, and CA literature. Furthermore, we contribute to the broader domain on security incident communication. Considering data breaches as specific security incidents [10], the application of the developed CA can be transferred to more general security incidents (e.g., ransomware or DDOS attack). Thus, providing a prospective field of action for the management of security incidents and their responses.

2 First Iteration: Design Requirements

Table 1. Examples of two actual data breach notifications

MR#	Notification of Bombas [8]	Notification of British Airways [9]
MR1	*Dear [Customer Name]* *Sincerely, [Signature] – Co-Founder, CEO*	*We are investigating…* *British Airways continues…*
MR2	*Malware in the code of the e-commerce platform…41.000 customers who made a credit card purchase*	*...theft of customer data from our website and our mobile app…relates to customer bookings made or changed between…*
MR3	*Data accessed may have included personal information such as name, address, and credit card information…For further information and assistance, please contact…*	*Name, addresses, and all bank card details were all at risk.*
MR4	*Since the breach, we have implemented additional security measures…We advise you to remain vigilant*	*…reported the data theft to the Information Commissioner…We recommend you contact your bank and follow their recommended advice.*
MR5	*…we are offering you free identity monitoring for two years…please accept our apologies*	*We have contacted all affected customers to say sorry…Any customer…will be reimbursed… Further, we will offer a 12 month credit rating monitoring*

Building on the outlined problem statement, we define our problem space [20] as: *information and communication related assistance in the data breach response process.* We aimed to derive MR that are context-insensitive and response strategy agnostic. This corresponds to the idea of defining a "class of goals" [18] rather than specific contextual challenges. The derived MR are presented with reference to citations of two actual data breach notification (see Table 1).

Issuing a notification constitutes the first point of contact with a customer after a data breach. Accordingly, companies must approach their customers and inform them about the data breach. Properly addressing the affected target group plays a vital role in this process, as addressing the large number of stakeholders affected induces a range of different communication types. Moreover, relations between the company and its customers must be considered: How is the customer viewed and how is he or she associated with the company? This influences not only the form of speech and formality, but also the nature of text, shaping the way of communicating. This also include the assignment of authors to the notification (see Table 1).

MR1: Integrating user-specific information on the company-customer relationship. To fulfill an individual and adequate addressing of the affected stakeholders, the CDBA must be able to dynamically identify the customer group it is interacting with.

Table 1 highlights that both companies recall the malicious activities led to a data breach. Thus, framing the experienced data breach and introducing the customer to the setting. An overview of what happened during the data breach is supplied. While the information specified differs between companies, it may include the date of the attack, and of the recovery [2]. Moreover, the reason for the data breach is usually given (e.g., malware, and data theft) as well as general information about the data breach (e.g., number of affected individuals). Thus, the aim is to ensure that the customer has a broad overview about the data breach.

MR2: Providing a general overview and key information about the data breach. The CDBA must be able to provide a summary of the breach in which various characteristics of it are introduced.

Customers are not expected to know which notification they have received or, in the case of a CDBA, why they are interacting with it. Therefore, establishing a basis for customer comprehension becomes crucial. While providing overarching information is of relevance, customers mostly need to know which of their information is (not) affected by the data breach. Thus, the severity of a data breach, which is mostly determined by the type of information disclosed, is conveyed to the customer. This description includes the company's knowledge about details of the breached information (e.g., to what extent, see Table 1). Furthermore, the provision of additional customer-specific information is usually provided by a hotline. In the case of Bombas, customers can request more detailed information and ask about specific damages. We conclude in the context of a CDBA:

MR3: Provision of customer-specific information. The CDBA must be able to deliver specific information to the customer. All information that may be relevant, including customer-specific data, must be presented on an ad-hoc basis.

To respond to data breaches, companies employ a variety of measures which they also report in data breach notifications. Table 1 demonstrates that Bombas has already implemented security measures for a recurrence of a similar data breach. Providing this information is vital in educating customers about what a company has done and what it intends to do. Customers can further be educated on what actions to follow to minimize the damage. Both companies in Table 1 illustrate this by informing affected customers how they can respond and be adequately prepared, specifically regarding financial misuse.

MR4: Details of a company's response activities and customer recommendations. A CDBA must be able to explain the range of actions taken before, during, and after the data breach. Additionally, customer recommendations should be supplied.

We have already highlighted the relevance of strategy components in the context of communicative data breach responses (e.g. apology [3]). Strategy components are decisions to be used tactically, which depend on various boundary conditions [6]. An static incorporation of specific components in a CDBA, e.g. compensation, would contradict their strategic character [3]. However, their influence on the customer as well as on the company is critical. For instance, in addition to compensating customers through an identity monitoring service, Bombas also utilizes an explicit apology to customers. British Airways additionally offers a full refund (see Table 1). Accordingly, CDBAs must be able to address these components dynamically.

MR5: Integration of strategy components. While considering all other MRs, a CDBA must incorporate the company's context-specific strategy component applied.

3 Next Iterations

As the first step of our iterative DSR endeavor, we completed the derivation of the MR. To transform these into design principles, we intend to engage with communication experts in the security domain who are experienced in the specificity of data breach communication. We will then derive testable propositions (TP) and evaluate a first prototypical artefact in a scenario-based experiment (virtual data breach setting). This artifact will be implemented as a corporate CDBA so that customer-specific information can be dynamically extracted from internal systems. We aim to compare the customer response to a data breach notification versus the response to a CDBA. For this comparison, a vignette study (1x1) is conducted. Building on these findings, we will adapt the design principles and refine the artifact. By drawing on restorative justice, this research aims not only to dissolve the traditional boundaries of closed communication in the field of IT security, but also to explore new application areas for CAs. In addition to the literature on data breaches, our research also contributes to the field of security incidents. Overall security incident response management (e.g., ransomware or DDOS attacks) can leverage the developed CA and the results of the study.

References

1. Nintendo: Data Breach Notification, https://www.nintendo.co.jp/support/information/2020/0424.html, last accessed 2022/03/26.
2. Foerderer, J., Schuetz, S.W., Foerderer, J., Schuetz, S.W.: Data Breach Announcements and Stock Market Reactions : A Matter of Timing ? Data Breach Announcements and Stock Market Reactions : A Matter of Timing ? (2022).
3. Gwebu, K.L., Wang, J., Wang, L.: The Role of Corporate Reputation and Crisis Response Strategies in Data Breach Management. J. Manag. Inf. Syst. 35, 683–714 (2018).
4. Bansal, G., Zahedi, F.M.: Trust violation and repair: The information privacy perspective. Decis. Support Syst. 71, 62–77 (2015).
5. Martin, K.D., Borah, A., Palmatier, R.W.: Data privacy: Effects on customer and firm performance. J. Mark. 81, 36–58 (2017).
6. Goode, S., Hoehle, H., Venkatesh, V., Brown, S.A.: User Compensation as a Data Breach Recovery Action: An Investigation of the Sony PlayStation Network Breach. MIS Q. 41, 703–727 (2017).
7. Wang, X., Wang, X., Liu, Z., Chang, W., Hou, Y., Zhao, Z.: Too generous to be fair? Experiments on the interplay of what, when, and how in data breach recovery of the hotel industry. Tour. Manag. 88, 104420 (2022).
8. Bombas: Bombas Data Breach Notification, https://oag.ca.gov/system/files/Bombas Ad r4prf %28002%29_0.pdf.
9. British Airways: Customer data theft, https://web.archive.org/web/20200316153540/https://www.britishairways.com/en-gb/information/incident/data-theft/latest-information, last accessed 2020/08/16.
10. Janakiraman, R., Lim, J.H., Rishika, R.: The Effect of a Data Breach Announcement on Customer Behavior: Evidence from a Multichannel Retailer. J. Mark. 82, 85–105 (2018).
11. Verstraete, M., Zarsky, T.: Optimizing breach notification. Univ. Ill. Law Rev. 2021, 803–864 (2021).
12. Van Ness, D.W.: Crime and its victims. Intervarsity Press (1986).
13. Braithwaite, J.: Restorative Justice: Assessing Optimistic and Pessimistic Accounts. Crime and Justice. 25, 1–127 (1999).
14. Wenzel, M., Okimoto, T.G., Cameron, K.: Do Retributive and Restorative Justice Processes Address Different Symbolic Concerns? Crit. Criminol. 20, 25–44 (2012).
15. Diederich, S., Brendel, A.B., Morana, S., Kolbe, L.: On the Design of and Interaction with Conversational Agents: An Organizing and Assessing Review of Human-Computer Interaction Research. J. Assoz. Inf. Syst. 23, 96–138 (2022).
16. Janssen, A., Passlick, J., Rodríguez Cardona, D., Breitner, M.H.: Virtual Assistance in Any Context: A Taxonomy of Design Elements for Domain-Specific Chatbots. Bus. Inf. Syst. Eng. 62, 211–225 (2020).
17. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS Q. 28, 75–105 (2004).
18. Walls, J.G., Widmeyer, G.R., Sawy, O.A. El: Building an Information System Design Theory for Vigilant EIS. Inf. Syst. Res. 3, 36–59 (1992).
19. Venable, J., Baskerville, R.: Eating our own Cooking: Toward a Design Science of Research Methods. Electron. J. Bus. Res. Methods. 10, 141–153 (2012).
20. Venable, J.R.: The Role of Theory and Theorising in Design Science Research. In: Proceedings of the 1st International Conference on Design Science in Information Systems and Technology (DESRIST 2006). pp. 1–18 (2006).