BRITISH JOURNAL of MANAGEMENT
BRITISH ACADEMY OF MANAGEMENT

British Journal of Management, Vol. 00, 1–18 (2020)
DOI: 10.1111/1467-8551.12427

Towards an Understanding of Privacy Management Architecture in Big Data: An Experimental Research

Nick Hajli,[1] Farid Shirazi,[2] Mina Tajvidi[3] and Nurul Huda[4]

[1]School of Management, Swansea University, Swansea, UK, [2]Ted Rogers School of Information Technology Management, Ryerson University, Toronto, Canada, [3]Newcastle University Business School, Newcastle University, London, UK, and [4]Department of Information & Communication Engineering Technology, Centennial College
Corresponding author email: Nick.hajli@swansea.ac.uk

Big data analytics provide valuable information allowing organizations to gain insights that grant them a competitive advantage in the market. However, it also provides access to data that compromise people's privacy. The development of sophisticated technologies for data analysis has resulted in a growing concern around privacy management in big data. While many sites (e.g. Facebook) require the user to provide personal information to access their services, others (e.g. Google search) can automatically capture or trace user activities and use that data to acquire personal demographic information. Therefore, Internet users are – willingly or unwillingly – constantly disclosing sensitive personal information. In addition, users do not get a complete picture of how their personal information is disseminated online. In this paper, we investigate information privacy through an experiment using large-scale disclosure of personal web activity data to track fragments of personal information released over a period of time. This experiment gives a clear picture of the potential privacy losses of individual users based on released personal information and activities at different websites. By devising an enterprise architecture using a privacy-by-design framework, this study provides a useful guide to addressing the managerial challenges of privacy management.

Introduction

'Big data' refers to an assortment of huge data sets that provide an avenue of analysis and yield associated patterns and trends. The amount of data being created and stored on a global level is almost inconceivable and is constantly growing. Big data is not just big: it is also diverse, and encompasses many different types of data, both locally stored and streaming (Chari, 2014). The data are being generated from physical space and are stored and processed in cyberspace; they include life logs (e.g. location, activities, communication history), online social network (OSN) data, public statistics and metadata, among other types. This data can be analysed or mined for valuable information, to reveal pertinent trends, and to enable organizational decision-making processes. Big data analytics empower organizations with additional resources that grant them a competitive advantage in the market (Dubey *et al.*, 2019). Privacy concerns can engender a breach of trust or even cause financial harm to people whose data are harvested in their encounters with organizations seeking a competitive advantage in the market (Herschel and Miori, 2017).

Much research has been devoted to the importance of big data analytics in creating value for businesses (Constantiou and Kallinikos, 2015; Mikalef *et al.*, 2019) and its role as a good source of organizational capability (Srinivasan and Swink, 2018), empowering firms to attain a competitive advantage (Akter *et al.*, 2016; Gupta and George, 2016). Although these studies emphasize the value of big data analytics in creating competitive advantage in the market, others argue that firms using big data open themselves to a variety of ethical dilemmas (Günther *et al.*, 2017), such as leaks of personal information; organizations that fail to properly manage such incidents may suffer consequences that affect their capabilities and indeed their competitive advantage. Currently, big data analytics is raising concerns over ethical issues of acquisition and usage of personal data. Data are generated and transmitted across the Internet either through an automated data-capturing system (e.g. a location-aware service) or through direct interactions with human beings, such as entering user information into an online form.

Many sites incentivize their clients to create accounts by rewarding them with a richer interaction experience. Some operational risk events occur frequently over a period of years due to the frequent social interaction among users in online environments. For instance, Equifax, a leading credit management services company in the USA, announced that personal identity and information (e.g. social security numbers, dates of birth and driver's licence information) of 143,000,000 US customers had been leaked after a massive hack. The news that customers' financial information and passwords had been stolen had serious consequences for the company; Equifax's share price fell 35% in a single week (The Buzz, 2017). Other high-profile hacks in recent years have included JPMorgan Chase in 2014 (Wired, 2015), Ashley Madison in 2015 (The Washington Post, 2017), Deloitte in 2017 (The Guardian, 2017) and Facebook in 2018 (The New York Times, 2018). In these breaches, hackers were able to access the personal information of customers and employees ranging from usernames and passwords to real names, addresses, phone numbers, email addresses, email content, relationship status, religion, date of birth, workplaces, search activity, banking data, credit card transactions, recent location check-ins and even secret sexual fantasies.

It is widely acknowledged that privacy protection is an important issue in online services (Cohen, 2018). However, it is currently a difficult task to keep track of the specific pieces of information a user has shared either knowingly, as through direct interaction with websites, or unknowingly (e.g. as a side-effect of explicit web activity) (Elahi, d'Aquin and Motta, 2010). Although previous research (Batistič and van der Laken, 2019) highlights the importance of privacy in big data analytics, we argue that debate, as currently framed, proceeds from an incomplete understanding of the problem. To wit: not much has been written about the practical ethical concerns surrounding big data – that is, about the mechanisms by which users' online activities can lead to the dissemination and misappropriation of personal information.

To address these challenges, the aim of this study is to analyse the extent to which today's big data settings have increased the potential for privacy harms in the form of loss of personal information. It is important to note here that this study is limited to analysis of data extracted from websites. This being the case, other sources of data – such as cloud-based applications (e.g. Dropbox), sensor data from Internet of Things (IoT) devices and mobile applications – are excluded from this study.

To address these challenges, we formulate the following questions:

1. What factors lead to direct leakage of information on popular websites?
2. How can we tackle the shortcomings of existing privacy measures through enterprise architecture (EA) design?

To answer our research questions, we focus on addressing the leakage and potential linkage of personal information from different sites. To date, few studies have utilized a theory-focused approach to addressing the challenges of big data research. In this study, we use a theoretical drawing from the resource-based view, dynamic capabilities and privacy concerns to address the research questions. We also look at a broad array of sites in various categories, based on their Alexa categories. This taxonomy derives from Alexa, a data analysis tool that aims to provide analysis for global data from millions of Internet users (www.alexa.com). We examine the extent of the direct leakage of private information as a result of typical user actions on these sites. We define 'direct leakage' as the loss of privacy through information flows (Li *et al.*, 2015). We then consider exactly which subsets of private information the aggregators receive from these websites. System monitoring and tracking data disclosure can make users aware of potential privacy loss. We explore the potential for aggregators to link various pieces of information they receive via GUIDs (such as user IDs from these sites) or via online activities. Finally, using those theoretical lenses, we fill the current research gap in big data analytics with regard to concern over ethical issues of acquisition and usage of personal data by developing an EA using the privacy-by-design (PbD) framework to address the shortcomings of existing privacy measures. PbD is a holistic approach for applying information technology (IT) to business practices, processes, physical design and network infrastructure (Stallings, 2020). Under this approach, privacy is embedded into the system's design and implementation. The PbD approach emphasizes the four privacy protection practices: purpose specification, collection limitation, data minimization and discourse limitation (Bennett, 2011; Bennett and Raab, 2006; Clarke, 2000; Wright and Raab, 2014).

The contribution of this research, in short, is that we utilize a theory-based focus to address the challenges of privacy in big data research through experimental research. In addition, the present study serves as a point of intersection in the discourse between investigation of strategic issues and operational implementations in the context of big data research (Batistič and van der Laken, 2019). For example, in the contexts of both operational management and behavioural theories, the developed EA framework offers opportunities for organizations to deal with ethical issues such as ensuring the protection of customers' privacy in the era of big data and information sharing.

While big data allows firms to rapidly capture, analyse and exploit information, it can also enable access to data that compromises an individual's privacy, either deliberately or inadvertently (Herschel and Miori, 2017). As such, continuously exploring and testing the ethical implications of analytical initiatives would allow organizations to establish their long-term strategic planning more firmly (Batistič and van der Laken, 2019). Knowledge drawn from the field of behavioural ethics will allow academics and managers to gain insight into improving the ethical environment in their organizations in a sustainable manner (De Cremer *et al.*, 2011). The behavioural approach that we advocate, as outlined by De Cremer *et al.* (2011), will stress the ethical and privacy content associated with the principal–agent models by arguing that the role of the individual is central to understanding corporate behaviour (Ullah *et al.*, 2019). This is in particular important given that previous research by Bell and Bryman (2007) also suggests that scholars in the management domain are more likely to experience conflicts of interest or affiliation bias, as their views often differ from those of their peers as to what constitutes ethical behaviour. More research is required to better understand the factors that influence an individual's moral behaviour within an organization, and thus ultimately to design better ethical infrastructure (Treviño, Weaver and Reynolds, 2006).

The results of our experimental research introduce the principles of a PbD framework based on EA. The PbD framework is designed to address the gaps in existing privacy frameworks and respond to dramatic changes wrought by big data, artificial intelligence (AI) and the IoT landscape. The findings represent an effort to address the gap in existing EAs by incorporating the PbD framework into EA design and implementation. Another theoretical contribution of the present research is to provide a solution to integrating privacy protection into EA. These innovations will help us to introduce our main contribution: a methodology for identifying categories of sites, determining a specific set of sites to study within each category, and then describing the ways in which the data were analysed – both for leakage of private information and for potential linkage to third parties.

The rest of this paper is organized as follows. The next section describes our study on the leakage of private information, and the effectiveness of enterprise system modelling in addressing the issue. The third section describes our research methods, while the fourth section discusses the results of our study. The fifth section examines the shortcomings of existing privacy protection measures and new schemes to combat the leakage and linkage of private information, as discussed in a case study. The final section presents a summary of our findings, discusses our theoretical contribution and its practical implications and suggests directions for future work.

Literature review and theoretical background

In the following subsection, we discuss using the resource-based theory in conjunction with dynamic capabilities to address the privacy issues associated with emergent technologies such as big data.

The resource-based view and dynamic capabilities

The resource-based theory views the organization as a collection of resources, positing that organizations use these resources to gain competitive advantage (McKelvie and Davidsson, 2009). Multiple studies highlight the value and impact of big data analytics on firms' performance (Akter *et al*., 2016; Aydiner *et al*., 2019; Chen, Preston and Swink, 2015; Dubey *et al*., 2019; Gupta and George, 2016). Most of these studies argue that organizations using big data analytics gain a competitive advantage in the market (Dubey *et al*., 2019; Srinivasan and Swink, 2018; Wamba *et al*., 2015) by using current data to forecast future events. One well-cited study on the resource-based view (Barney, 1991) explains how organizations can gain a competitive advantage by developing a package of strategic resources. IT capabilities such as big data analytics are among those valuable assets if combined with other organizational resources to create value and gain competitive advantage.

Other studies (Aydiner *et al*., 2019; Dubey *et al*., 2019) argue that the desired level of performance will not provide sufficient advantage if an organization fails to respond to external pressures in a timely fashion. The issues of information leakage and privacy are among those relevant external pressures which, if not effectively addressed, can prevent organizations from achieving the desired level of performance. The current research integrates the resource-based view and dynamic capabilities to develop an EA address the managerial challenges of privacy management.

Dynamic capabilities support organizations in obtaining competitive advantage (Eisenhardt and Martin, 2000; Zahra, Sapienza and Davidsson, 2006). The assumption is that organizations which have a good formulation of internal and external competencies carve out a better position for themselves than those that do not (McKelvie and Davidsson, 2009). Those processes that collect resources and integrate them to create new valuable assets allow the firm to gain competitive advantage in a rapidly changing market (Eisenhardt and Martin, 2000; Teece, Pisano and Shuen, 1997) – for instance, by using big data analysis to gain insight into the market or predict future trends. Knowledge creation is an example of such dynamic capability; big data analytics provide new insights based on current data, which helps develop new thinking in an organization.

The resource-based view and dynamic capabilities both emphasize the role of using organizational resources – including big data analytics – to create a competitive advantage. If such an advantage is to be sustainable, however, the organization must consider the ethical issues around using big data analytics, such as privacy concerns.

Privacy

Privacy has been studied for many years in almost all spheres of social science, including law, economics, sociology, political science, psychology, organizational behaviour, operational management, marketing and management information systems (Dinev *et al*., 2013).

Solove (2008) points out that in order to conceptualise privacy, we should view and understand it from its pluralistic context, rather than as a unitary common denominator – that is, instead of attempting to locate the common denominator of the freedoms comprising 'privacy', we should conceptualize it by focusing on the domains of disruption. In this framing, privacy does not have the same universal value across all contexts and cultures.

One of the earliest scholarly frameworks for privacy was introduced by Clarke (1997, 2000), who identified five different categories: personal data, communications, experience, behaviour and bodily privacy. Another privacy framework was Solove's (2008) privacy taxonomy, consisting of four main groups of activities (information collection, information processing, information dissemination and invasion), each introducing a variety of issues to be addressed. However, Finn, Wright and Friedewald (2013) argue that Solove's taxonomy focuses mainly on privacy problems rather than characterizing the types of privacy.

The PbD framework is designed to address the gaps in existing privacy frameworks and respond to dramatic changes in the ICT landscape – in particular, emergent technologies such as big data, AI, the IoT and decentralized networks. In essence, PbD grew out of earlier frameworks, including Clarke's and Solove's, driven by the need for privacy engineering within the context of EA. Dinev *et al*. (2013) point out that the notions of privacy in the academic literature are mainly discipline-specific; as such, the concepts, definitions and relationships are inconsistent, and are neither fully developed nor empirically validated. For example, legal scholars define privacy as a 'right' or 'entitlement', while other disciplines – including philosophy and psychology – define it as a 'state of limited access or isolation'; still others, particularly the social sciences and information systems, use 'control' as a definition of privacy (Dinev *et al.*, 2013).

As regards IT, Culnan and Williams (2009) argue that information privacy is a multidimensional concept in which privacy problems result from the subsequent storage, analysis, use or sharing of information. Agre and Rotenberg (1998) point out that databases as a means of organized information storage have historically been designed on the assumption that data records can be traced back to the subjects they represent; in practical terms, though, such tracing is often unnecessary, and the stored data can be compromised. This information tracing, accompanied by information reuse deployed by many organizations and unauthorized access (e.g. employees viewing personal information they are not authorized to view), can potentially threaten an individual's ability to maintain a condition of limited access to his/her personal information (Culnan and Williams, 2009). The issue is exacerbated by the autonomous collection of private information (Nunan and Di Domenico, 2013), conducted independently of human activity – for instance, by automated processes or robots.

Today, technologies such as cloud computing, big data analytics, the IoT and decentralized technology environments contribute to a different organizational privacy problem: data breaches (Culnan, Foxman and Ray, 2008). Dattner *et al.* (2019) argue that as technology advances, big data and AI are able to determine 'proxy' variables for private, personal attributes with increased accuracy. For example, as mentioned by Chamorro-Premuzic, Polli and Dattner (2019), AI has disrupted every area of our lives, from business process design and online shopping experiences to the personalized recommendations that channels like YouTube and Netflix use to market their latest content; but AI is still in its infancy. While these novel tools are disrupting the recruitment and assessment space, they raise questions about their accuracy and the ethical, legal and privacy implications that they introduce. This being the case, the notion of privacy engineering (Stallings, 2020) has emerged in the fields of IT and management. Privacy engineering involves taking account of privacy during the lifecycle of ICT systems, such that privacy is and remains an integral part of their design and functionality (Stallings, 2020). In fact, privacy engineering constitutes a mapping and implementation of PbD, according to the European Data Protection Supervisory agency (EDPS, 2018).

Information leakage. One of the most frequently occurring issues in the business world is information leakage – the unauthorized transmission of data, electronically or physically, from within an organization to an external destination or recipient (Zhang *et al.*, 2012). Policy-makers in many countries have formulated laws and regulations to standardize the ways firms can use the data collected on their websites and require them to develop big data ethics policies (Boyd and Crawford, 2012; Herschel and Miori, 2017). Although this legislation aims to protect user rights and prevent illegal discrimination, mechanisms to monitor these information exchanges (e.g. web browser histories) are still not commonplace or are generally very limited (O'Leary, 2016).

Enterprise architecture modelling has been recognized as an effective approach to addressing the challenge of standardizing privacy and security protocols; implementing EA modelling can help companies accurately assess the impact of various components across EA and analyse the corresponding changes or ripple effects (Dam, Lê and Ghose, 2016).

Enterprise architecture

Tamm *et al.* (2011) define EA as the high-level representation of a business's processes and systems and the interlinking between them. The purpose of EA is to establish a company-wide culture that reflects the organization's values, norms and principles (Aier, 2014; Foorthuis *et al.*, 2016; Proper and Greefhorst, 2010; Ross, Weill and Robertson, 2006). However, the term is currently associated with the structure and functionality of an organization that diverges from its original purpose (Lankhorst, 2009; McGovern *et al.*, 2004; Owen and Raj, 2003). Enterprise architecture is composed of frameworks (e.g. Zachman, TOGAF) and implementation methodologies that support the development of models and IT infrastructure for the enterprise (Nikpay *et al.*, 2017; Rouhani *et al.*, 2013, 2015; Sessions, 2007; Shah and El Kourdi, 2007). Companies follow specific frameworks to successfully collect valuable data and use implementation methodologies to integrate and align business strategies with IT. However, the framework and methodologies a company chooses depend strongly on its goals.

Enterprise architecture methodologies. There are a variety of approaches to EA. First, Quartel *et al.* (2011) shed light on methods like *The Open Group Architecture Framework* (TOGAF), and link them to the significance of requirement engineering within the context of EA. They propose a language to support the modelling of motivation.

Second, as Tamm *et al.* (2011) suggest, there are several ways that EA offers organizational benefits as defined in the EA benefits model: privacy, organizational alignment, resource portfolio optimization and resource complementarity. The choice of operating platform must factor in the future objectives and goals of an organization, and more focus is needed on the ways in which EA leads to organizational benefits (Tamm *et al.*, 2011). This would enable a greater integration of privacy with EA.

Third, other research focuses on the role of EA and the management of change within complex organizations. This is done through a focus on the role of motivation within EA (Yu, Strohmaier and Deng, 2006), and is further linked to privacy and security. As EA is meant to extensively depict the core components and relationships in a given enterprise, it is essential for change management (Yu, Strohmaier and Deng, 2006). Currently, most EA modelling does not place the necessary focus on motivation, and Yu and colleagues suggest two intentional modelling languages – Entity Relationship (ER) and Unified Modelling Language (UML) – as means to address this shortcoming in the EA construction processes.

Fourth, there is focus on the use of EA in the propagation of strategy and process changes, the management of IT/business and the consistency of business transformation (Fischer, Aier and Winter, 2007). This approach sheds light on the weaknesses of existing EA models and highlights the increasing acceptance of EA in change management and for IT/business alignment.

Enterprise architecture also has a need for requirements modelling and for modelling support (Engelsman *et al.*, 2011). This ties in with the focus on requirements engineering, privacy, goal modelling, stakeholder concerns and the alignment of business and IT.

A privacy-by-design view

As proposed by Cavoukian (2009), PbD addresses the growing issue of privacy on larger networks. Privacy-by-design provides a perspective on privacy and how regular policies and frameworks alone cannot ensure compliance and security. It covers aspects of information systems, business practices and networked infrastructure, and offers seven fundamental principles to serve as foundations for practices to safeguard one's personal information. These principles include stipulations that a PbD framework must be proactive, not reactive; be preventative, not remedial; and take a strategic view, rather than respond to issues after they occur (e.g. crash recovery).[1]

The main aim in deploying PbD principles in this study is to set up a privacy management protocol that adheres to the principles of privacy as a default setting. Purpose specification refers to the purposes for collecting, utilizing, retaining and disclosing data, which should be communicated to users before the data are collected. Collection limitation describes the fairness and legality of data collection, while data minimization defines the principle of collecting only the strict minimum of personal information necessary to the task at hand. Disclosure limitation refers to limiting the use, retention and disclosure of personal identifiable information (PII), such as names, social insurance or security numbers, passport numbers, driver's licence numbers, street addresses, telephone numbers, medical data such as X-rays, biometric data such as images or fingerprints, vehicle registration numbers and other information linked or linkable to one of the above identifiers, such as date of birth, place of birth, race, religion, employment information or financial information (NIST, 2010).

Each individual should be empowered to grant or withhold consent for the collection, use or disclosure of PII. In this context, the inclusion of PbD within an EA allows for greater efficiency and integration of organizational processes.

[1] For more information about the seven foundational principles of PbD, please visit: https://www.ipc.on.ca/wp-content/uploads/resources/7foundationalprinciples.pdf

Research methods

The methodology of this study is experimental research. Experimental research has shown successful results in the information systems and management fields, among others (Loch and Wu, 2008; Tsai *et al.*, 2011; Venkatraman and Zaheer, 1990). Using an experimental study design, we have created a personal analytics dashboard to highlight the need to implement the principles of PbD in EA design to mitigate the privacy problems associated with data disclosure when surfing the Internet. The goal of this experiment is to help managers and decision-makers improve the design of their infrastructure networks so as to be better prepared to tackle the growing issues of cyber-attacks worldwide. We support our experimental design with a case study, as detailed below.

Categories and sites for study

Many types of websites encourage users to create accounts; doing so is often a prerequisite for job-search or dating sites. Other types of websites allow registered users to upload content, while restricting unregistered users to browsing content. Registered users can post reviews and comments, personalize webpages, participate in contests, save their purchase information, receive electronic newsletters and gain access to restricted site content. Although only a relative handful of users may value these functions enough to register, for high-volume sites that small fraction still represents a large number of users. In our study, we used Alexa's (www.alexa.com) categories and subcategories, including sites with a significant number of registered users that allowed for registration without any need for credit card information. Alexa categories have been defined in accordance with analysis of global data traffic.

Using self-reported numbers and published reports about each site, we set a threshold of at least 100,000 registered users (on most sites, this number was in the millions). Sites for which we were unable to find registration numbers were included, provided they were popular sites in a category where other sites had evidence of significant registration numbers.

To ensure that we had the most popular sites, we began with the top-ranked site in each Alexa category (or subcategory) and worked down the list, checking for the above criteria until we reached a target of 10 sites within each category. Categories that did not yield at least 10 sites were dropped. We identified 10 categories and subcategories for study (out of 17 Alexa categories): Adult, Arts, Business, Computers, Games, Health, Home, Kids and Teens, News, Recreation, Reference, Regional, Science, Shopping, Society, Sports and World.

Additionally, we examined the OSN and Health categories, given that users often supply potentially sensitive information to such sites. Search terms used or pages viewed could indicate interest in sensitive medical conditions, and the availability of such information to third parties could result in its being linked with other private information about a user. We used a similar methodology for determining 10 sites from the OSN and Health categories, although in this case we relaxed the requirement of requiring user registration, as private information could be leaked from these sites even without explicit user identification. For the purposes of our study, we established an account if the site in question offered a provision for registration. In doing so, we made sure that there were no ethical problems associated with lurking, or with engaging other users in online conversation without their consent.

Data capture methodology

We followed the data capture methodology depicted in Figure 1. We first captured HTTP requests and responses from our web browser using the Fiddler web proxy and examined them for visibly transmitted private information. Encrypted information or information transmitted over Secure Socket Layer (SSL) could not be observed, except for a tiny fraction; as a result, the transmission of data over SSL did not lead to observable data loss.

The initial steps for testing each site consisted of creating an account, confirming a verification email message if needed and viewing/editing the user profile on the site. Some sites provide an opportunity to 'Remember Me' on login; this was selected when available, as it allowed the site in question to then store private information (e.g. in cookies) and subsequently potentially leak it to a third party. A number of sites in our study allow users to create an account and sign in via an existing third-party account, such as Facebook,

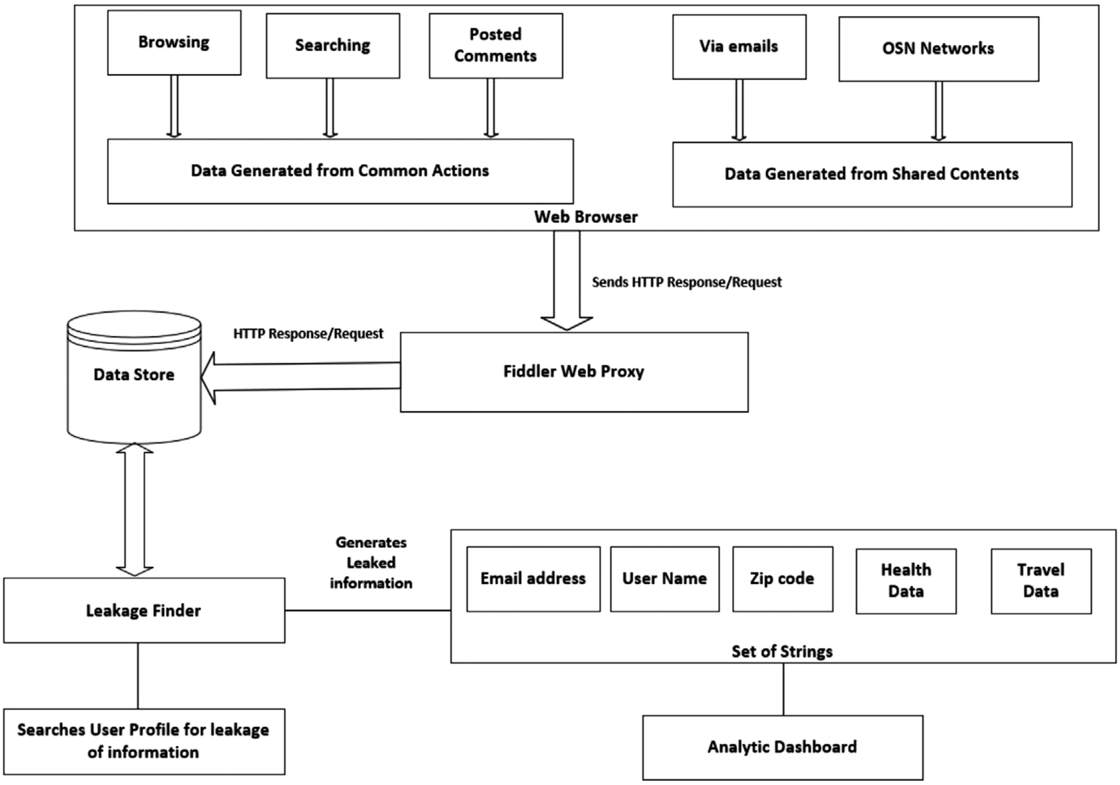


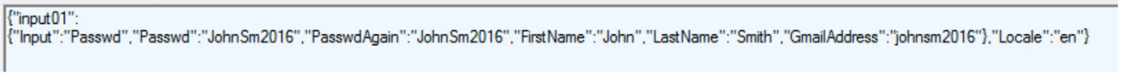
Figure 1. Data capture methodology

Google or Twitter. In cases where we could establish an account directly with the first-party site, we always chose that option.

The set of actions for which we tested varied according to each site's category and were tailored to the site's available features. Where feasible, we included actions that used features only available to registered users. In some categories (e.g. Arts and News), we included actions available to all users. Many of the sites featured a set of common actions: browsing, appropriate searches and posting comments or reviews on the site's content pages. Beyond category-specific actions, many sites also provide opportunities to share content with friends via email, or to connect with social networking sites. When available, we shared content with sites in our OSN category and emailed articles to 'friends'. Xu *et al*. (2019) argue that the privacy control mechanisms in current OSNs are a one-way control mechanism in that they only impose restrictions on users who want to access others' data; there is no strict restriction on users who post data that may unintentionally violate other users' privacy.

We searched the gathered HTTP request/response (and POST) data for each site for leakage of user IDs, usernames and pieces of private information to third-party servers. We generated a set of strings extracted from a user's profile that might be leaked to a third party. The set included all strings that users are required to enter into their profile at the time of account creation, such as email addresses, names and ZIP codes. We also included other strings that users typically add to their profile, and which might contain sensitive information. Beyond profile data, we sent search queries to Health and Travel websites (in the form of travel dates and cities), adding to the collection of search strings.

Using the Fiddler web proxy, we processed the resulting data by systematically looking for leakage in the HTTP headers, and manually eliminated false positive matches (e.g. a long string containing ZIP codes). When a leak was identified, we recorded the leaked information, manner of leakage and the third-party recipient(s). It is important to note that we only reported leakage that we directly observed; our results thus constitute a



```
{"input01":
{"Input":"Passwd","Passwd":"JohnSm2016","PasswdAgain":"JohnSm2016","FirstName":"John","LastName":"Smith","GmailAddress":"johnsm2016"},"Locale":"en"}
```

Figure 2. Disclosing personal information (name, email, password) when creating a Gmail account [Colour figure can be viewed at wileyonlinelibrary.com]

conservative estimate of the extent of leakage. We may not have observed leakage to a third party for a number of reasons – for example, if we failed to test an action available on a given site that led to leakage, or if information was sent in a form we could not detect, or if the transmission was encrypted (e.g. sent over SSL).

We created a personal analytics dashboard to provide the end user with an interface visualizing data disclosure. These visualizations included information about the types of activities performed on online systems, the resources accessed, the location of the user, the time of usage and the browsers and operating systems used.

Results of the study

In this section, we will show examples of data leakage to third parties after enumerating common actions for registered users of sites. We explore leakage across site categories and conclude with an examination of the sensitivity and identifiability of the information leaked to third parties.

The series of actions that users perform on sites include mandatory actions, such as creating and logging into an account, and popular actions, such as editing profiles and searching. In some cases, an interaction might take the form of a sequence of actions. Below, we present actual examples highlighting when private information is leaked to an aggregator.

Account creation/confirmation. The first step is account creation, which for some sites requires responding to an account confirmation email. We observed leakage of private information during this process when the information is transmitted as part of the Request-URI of an HTTP GET request and then this Request-URI is contained in the Referrer header for subsequent requests of embedded objects from third parties. URI is the uniform resource identifier of the resource to which the request applies. Basically, URI is a blanket term applied to either uniform resource locators (URLs) or uniform resource names (URNs).

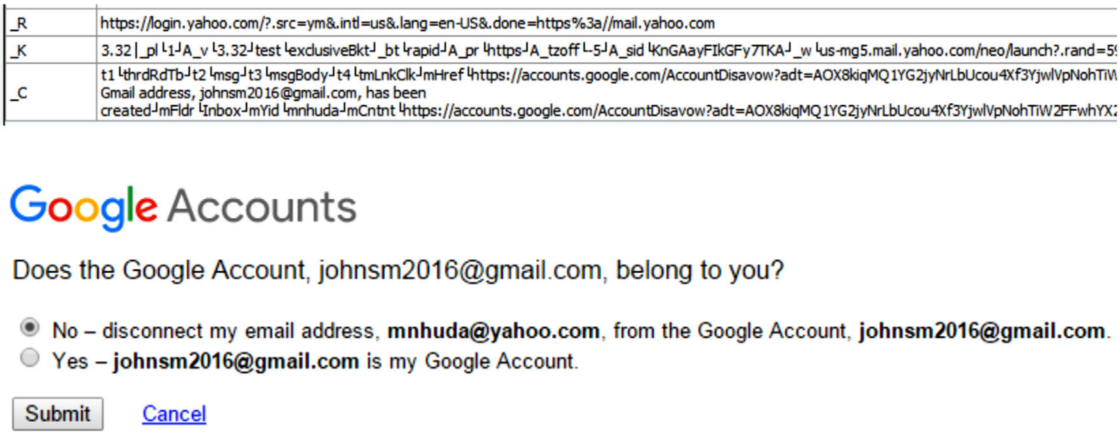
Figure 2 shows personal information transmitted to a Google server while the user was creating a Gmail account. It shows the linkage of another email address (a Yahoo Mail address) to the newly created Gmail account when the Yahoo Mail address was provided during the process of creating the Gmail account.

When the two email accounts are linked, all of the related data from both accounts can also be linked, as shown in Figure 3.

Searching for sensitive terms. Search terms are highly sensitive, and users expect them to stay entirely within the site where they are entered. Figure 4 shows an example wherein the search term 'Recovery from drug addiction' is sent to a google.ca server via another site.

It is important to note that although our experimental study was limited to HTTP protocol, we must also consider the security and privacy issues associated with HTTPS protocol using browser packages such as Mozilla Firefox or Google Chrome. As we know, many users share their workstations and personal computers with other users, including strangers, or they use computers in public spaces. In addition, many of us have experienced browser issues, such as a lack of response to commands and/or an unexpected browser crash. If a browser crashes and the user leaves the workstation, a second user can restore the crashed website and read the content of a previously opened window, even if it is related to a secure business email or a personal social media page; more importantly, the second user can change the settings of the original profile because the remote server assumes that it is dealing with an already-authenticated user. Figure 5 shows a snapshot of Firefox and Chrome crash reporter dialog boxes. By clicking on the 'Restart' or 'Restore' buttons, an adversarial actor can restore all previously open pages to their state before the crash.

As suggested by Tavani and Moor (2001), the individual control of personal information plays an important role in the management of privacy. In addition, sensitive personal information ought to remain private even if its owner is not in a position



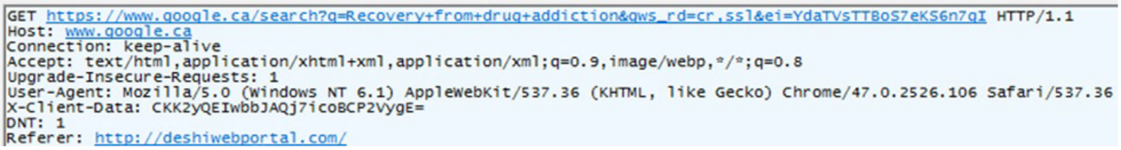
_R	https://login.yahoo.com/?.src=ym&.intl=us&.lang=en-US&.done=https%3a//mail.yahoo.com
_K	3.32 | _pl └1┘A_v └3.32┘test └exclusiveBkt┘_bt └rapid┘A_pr └https┘A_tzoff └-5┘A_sid └KnGAayFIkGFy7TKA┘_w └us-mg5.mail.yahoo.com/neo/launch?.rand=5
_C	t1 └thrdRdTb┘t2 └msg┘t3 └msgBody┘t4 └tmLnkClk┘mHref └https://accounts.google.com/AccountDisavow?adt=AOX8kiqMQ1YG2jyNrLbUcou4Xf3YjwlVpNohTiW Gmail address, johnsm2016@gmail.com, has been created┘mFldr └Inbox┘mYid └mnhuda┘mCntnt └https://accounts.google.com/AccountDisavow?adt=AOX8kiqMQ1YG2jyNrLbUcou4Xf3YjwlVpNohTiW2FFwhYX

Google Accounts

Does the Google Account, johnsm2016@gmail.com, belong to you?

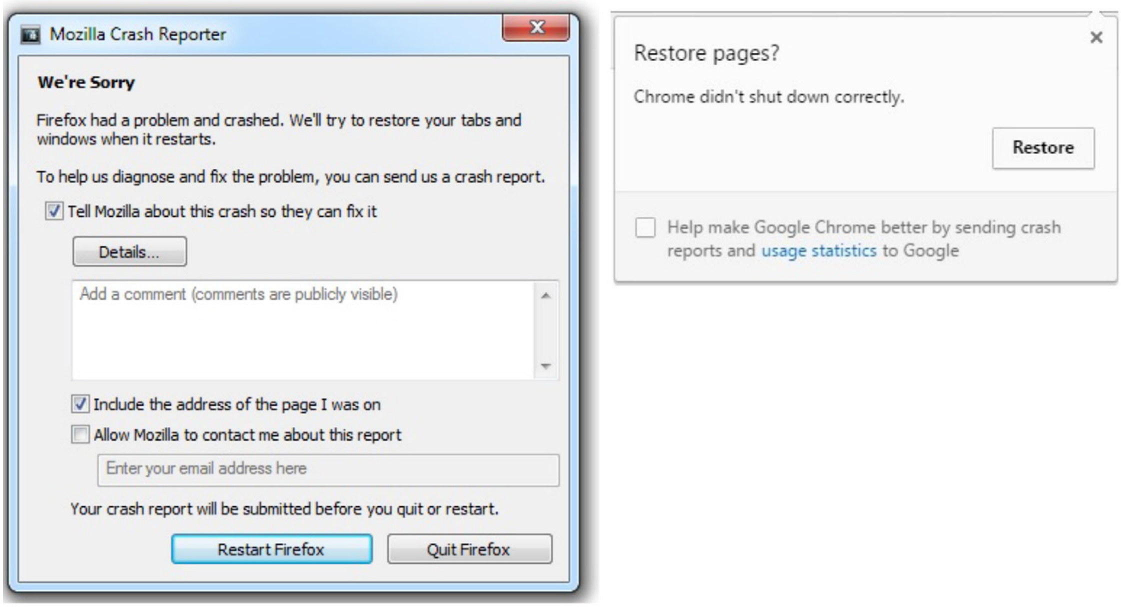
● No – disconnect my email address, **mnhuda@yahoo.com**, from the Google Account, **johnsm2016@gmail.com**.
○ Yes – **johnsm2016@gmail.com** is my Google Account.

Figure 3. Linking (and disclosing the linkage) of one email address to another [Colour figure can be viewed at wileyonlinelibrary.com]



```
GET https://www.google.ca/search?q=Recovery+from+drug+addiction&gws_rd=cr,ssl&ei=YdaTVsTTBoS7eKS6n7gI HTTP/1.1
Host: www.google.ca
Connection: keep-alive
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8
Upgrade-Insecure-Requests: 1
User-Agent: Mozilla/5.0 (Windows NT 6.1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/47.0.2526.106 Safari/537.36
X-Client-Data: CKK2yQEIwbbJAQj7icoBCP2VygE=
DNT: 1
Referer: http://deshiwebportal.com/
```

Figure 4. Search query revealing interest in 'Recovery from drug addiction', which could reveal a sensitive medical condition in the individual [Colour figure can be viewed at wileyonlinelibrary.com]



Mozilla Crash Reporter

We're Sorry

Firefox had a problem and crashed. We'll try to restore your tabs and windows when it restarts.

To help us diagnose and fix the problem, you can send us a crash report.

☑ Tell Mozilla about this crash so they can fix it

Add a comment (comments are publicly visible)

☑ Include the address of the page I was on
☐ Allow Mozilla to contact me about this report

Enter your email address here

Your crash report will be submitted before you quit or restart.

Restore pages?

Chrome didn't shut down correctly.

☐ Help make Google Chrome better by sending crash reports and usage statistics to Google

Figure 5. Typical crash report pages in Mozilla Firefox (left) and Chrome (right) [Colour figure can be viewed at wileyonlinelibrary.com]

to control it (Tavani and Moor, 2001). The above examples clearly demonstrate that current web applications do not adequately protect individual privacy (Carona *et al*., 2016). Collective efforts will be required to respond to today's privacy issues and the leak of private information. In particular, the principle of corporate social responsibility (CSR) holds that companies have a moral responsibility to stakeholders, which should be understood as a duty to protect information privacy

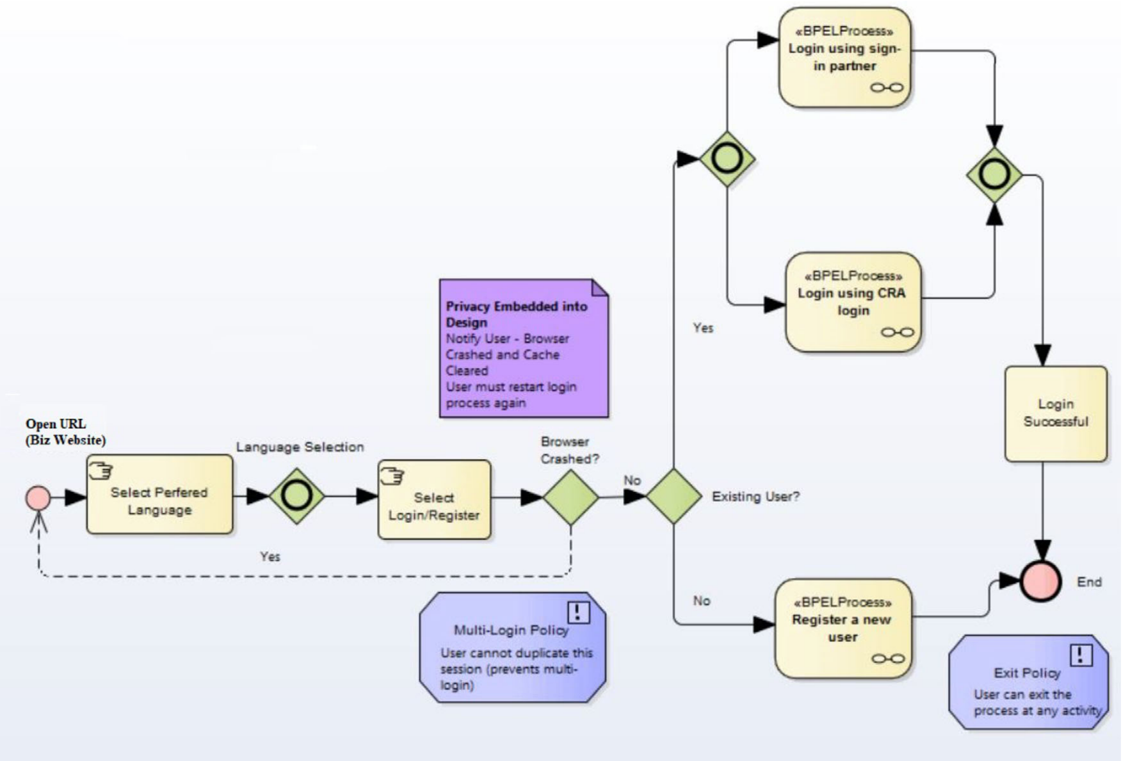


Figure 6. User login and the state of browser (crashed or not) [Colour figure can be viewed at wileyonlinelibrary.com]

and security. Social responsibility also requires a system that is equitable and efficient. Lindgreen and Swaen (2009) argue that while many consider CSR necessary for organizations to define their roles in society, some organizations struggle to apply social and ethical standards to their businesses (Lindgreen *et al*., 2009; Pinkston and Carroll, 1994). In addition, despite the well-accepted belief that CSR enables organizations to meet their stakeholder obligations, various unresolved issues remain (Lindgreen and Swaen, 2009). This study is an attempt to not only highlight the issues but also provide an EA solution for tackling them.

The following section responds to these issues by presenting a case study involving a secure website that allows multi-login through its partners' accounts (e.g. Google, Twitter, Facebook, or online banking accounts). We believe that the above issues are related to an EA-wide problem; this being the case, we offer a more robust design for building a privacy-preserving and trustworthy EA (Chen, Chiang and Storey, 2012) that adheres to the principles of privacy enumerated in the PbD framework (Cavoukian, 2009).

A multi-login case study

We conducted our project to test the vulnerability issues of a secure e-commerce website (Biz)[2] shortly after the site implemented enhanced security features. Currently, users can access their private financial information through previously created login credentials or through the use of third-party accounts, mainly online banking accounts. Using the EA tool ArchiMate 3.0 – which is designed around the TOGAF enterprise architecture and aligned with the PbD principles – we proposed the following new Biz settings.

This process is outlined in accordance with the diagram shown in Figure 6, which explains policy or principle objects and their content. When users arrive at the Biz link, they choose a language (English or French). After the preferred language has been selected, the user is prompted to enter his or her credentials. A successful login leads to the user being presented with his or her account

[2] Name restricted due to confidentiality.

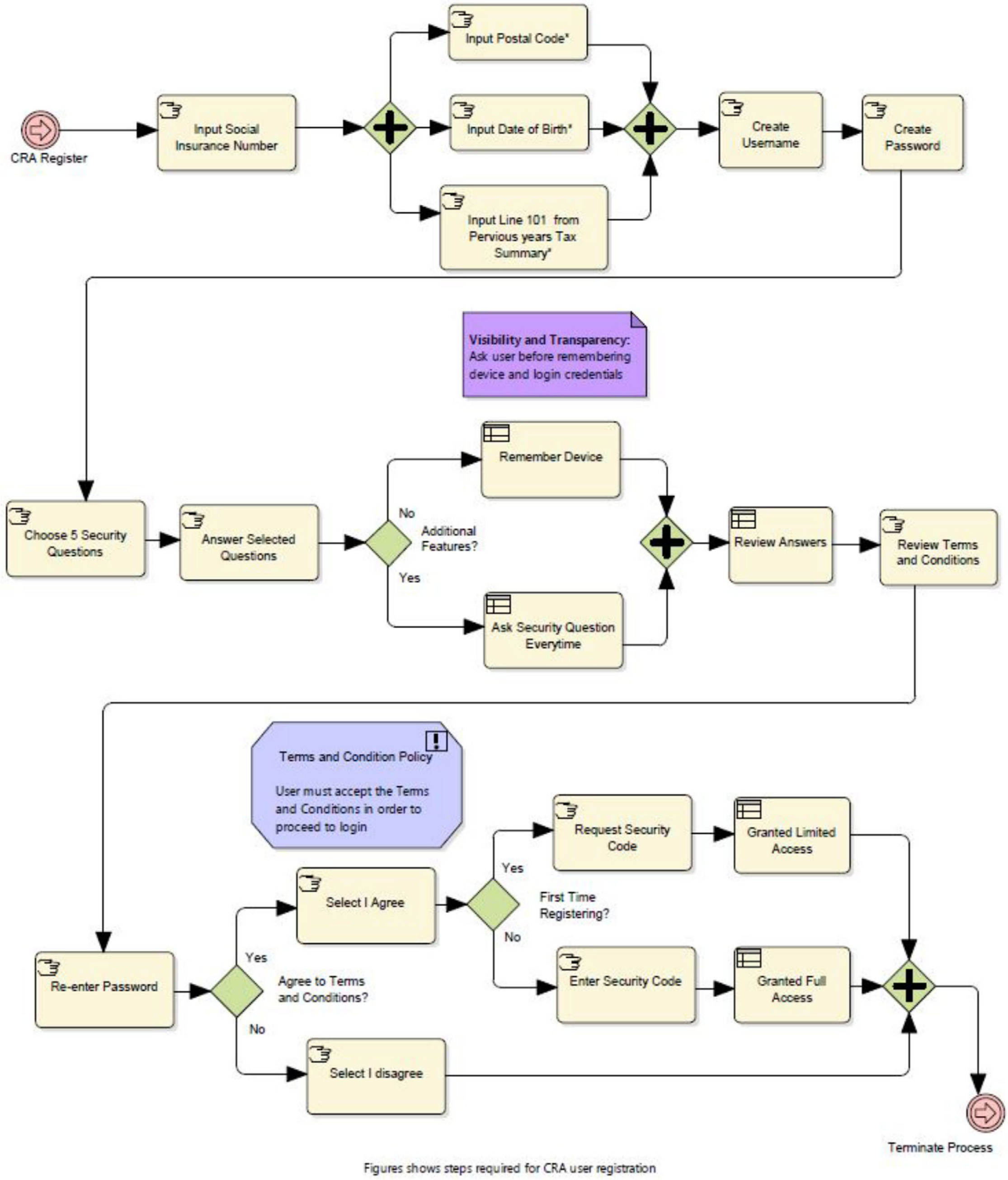
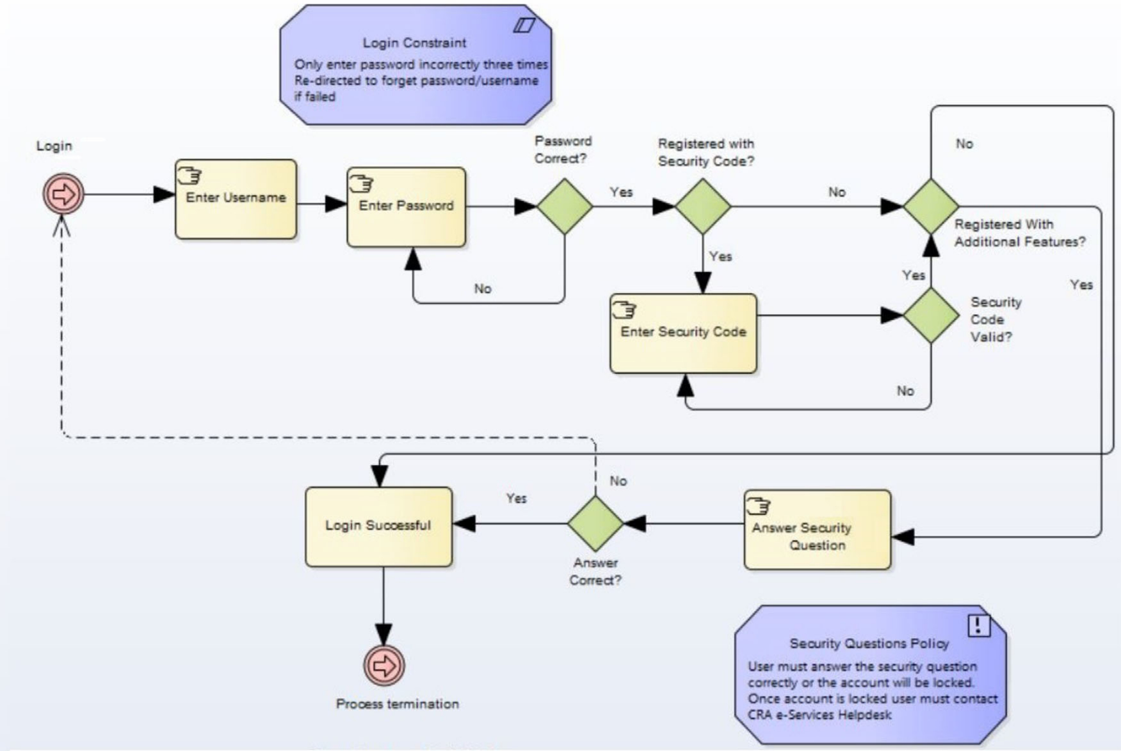

CRA Register
Input Social Insurance Number
Input Postal Code*
Input Date of Birth*
Input Line 101 from Pervious years Tax Summary*
Create Username
Create Password
Visibility and Transparency: Ask user before remembering device and login credentials
Choose 5 Security Questions
Answer Selected Questions
No
Additional Features?
Yes
Remember Device
Ask Security Question Everytime
Review Answers
Review Terms and Conditions
Terms and Condition Policy
User must accept the Terms and Conditions in order to proceed to login
Re-enter Password
Yes
Agree to Terms and Conditions?
No
Select I Agree
Select I disagree
Yes
First Time Registering?
No
Request Security Code
Enter Security Code
Granted Limited Access
Granted Full Access
Terminate Process
Figures shows steps required for CRA user registration

Figure 7. Steps required for Biz user registration [Colour figure can be viewed at wileyonlinelibrary.com]

screen. There is a possibility that at this point the browser crashes.

We tested a list of 11 supported financial institutions available on the Biz website. Only three financial institutions were able to recover from crash by prompting the user to restart the login process. In all other cases, the user was able to continue on the Biz server by restarting the crashed browser. (For confidentiality reasons, we exclude the names of the institutions in question.)

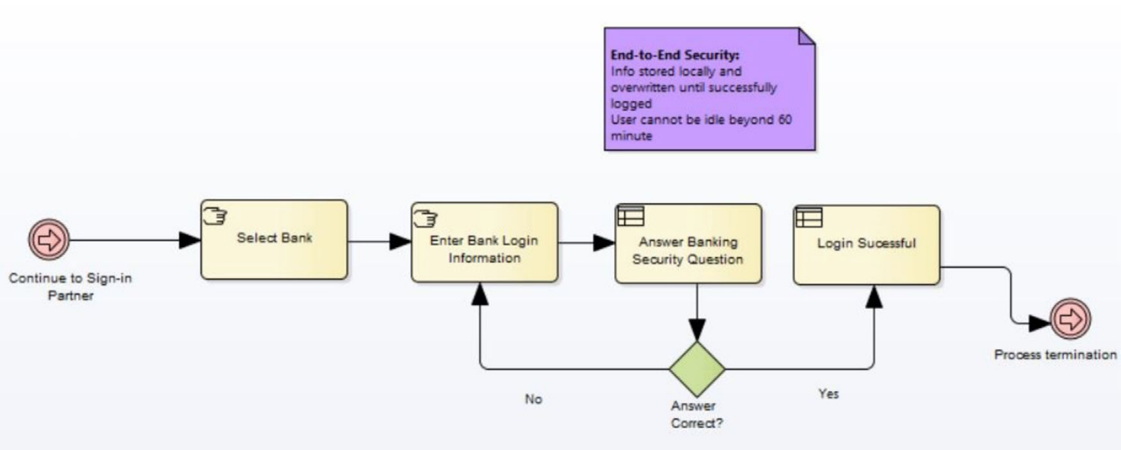
A multi-login policy would prevent bad actors from opportunistically acquiring sensitive data in this way. Our proposed model, as seen in Figures 7–9, prevents the user from performing



Login Constraint
Only enter password incorrectly three times
Re-directed to forget password/username if failed

Security Questions Policy
User must answer the security question correctly or the account will be locked.
Once account is locked user must contact CRA e-Services Helpdesk

Figure 8. Policies applied to the user login process [Colour figure can be viewed at wileyonlinelibrary.com]


Continue to Sign-in Partner
Select Bank
Enter Bank Login Information
Answer Banking Security Question
Login Sucessful
Answer Correct?
No
Yes
Process termination

End-to-End Security:
Info stored locally and overwritten until successfully logged
User cannot be idle beyond 60 minute

Figure 9. Login steps when using sign-in partners [Colour figure can be viewed at wileyonlinelibrary.com]

multiple logins by limiting duplication of a given session.

Figure 7 illustrates the steps required for Biz user registration, while Figure 8 applies our proposed policies and restrictions to the login process. Figure 9 illustrates the steps taken when a user decides to sign in using a Biz partner's account. Under the model depicted in these figures, privacy would be embedded within the design. Users would be notified in the event of a browser crash and be presented with their next steps – clearing the cache and restarting the login process.

An existing user has the option to continue either via a sign-in partner or through the Biz login (see Figure 9). If using a sign-in partner, the user would select from a list of banks and enter his or her banking information and the bank's security question. A correct answer would allow for a successful login; an incorrect answer would lead to a repeated prompt for banking information. The information is protected via end-to-end security. The Biz policy states that information is stored locally and overwritten until a successful login takes place. In addition, the user cannot be idle for a certain number of minutes.

If logging in via the Biz website, the user enters his or her username and password. If the user has been registered with a security code, and additional features like the Biz security question, then the user would be prompted for these options. The security question policy states that incorrect answers to the security question will lock the user's account. If the account gets locked, the only way to attain access would be by contacting the Biz helpdesk.

In contrast, a new user is prompted to enter his or her social security number, followed by date of birth, postal code and previously submitted tax information. After this, the user is prompted to create a username and password and to choose five security questions, inputting answers accordingly. The user instructs the website to either remember the device being used for the initial login, or to present security questions every time. This policy is built on visibility and transparency, and as a result, users are informed and give consent before the site remembers device and login credentials.

Our recommendation, as shown in Figure 7, is that after reaching a successful login screen, the user should have to review the terms and conditions, after which the user re-enters his or her password and affirms agreement to the terms. Non-acceptance of the terms terminates the process. Accepting the terms leads to either limited or full access, depending on whether the user has the security code. A first-time user would request a security code and be granted limited access. A previously existing user would enter his or her existing security code and be granted full access. Within this context, there are two policies that are of importance: the terms and conditions and the exit policy. The terms and conditions policy states that acceptance of the terms is required for the process to continue. The exit policy states that the user has the ability to exit the process at any point.

Including these policies would allow us to evaluate the changing structures in real time (Dhillon, Syed and Pedron, 2016) and help Biz address scenarios where a browser crash would allow the next user to access the previous user's system, as can happen when a user leaves his or her system unattended after a crash. A recovery of the crashed browser, as discussed above, would allow a new user on the system to access the previous user's home screen. With the inclusion of a browser crash policy, however, the user would be informed of the crash, and instructed as to the next necessary steps. In addition, there ought to be a prevention of multi-link logins, and a preference for Biz's own login process. This would decrease reliance on browsers to address browser crashes, and eliminate waiting on a third party for resolution.

Discussion of results, implications of research for managers and conclusion

Previous research has placed significant emphasis on the value of big data analytics as a means for organizations to gain insight into the market to gain a competitive edge (Akter *et al.*, 2016; Gupta and George, 2016; Srinivasan and Swink, 2018). Previous research (Herschel and Miori, 2017) highlights the issues associated with an individual's privacy raised by the use of big data. Occasionally, people are able to secure their identities anonymously, but this is not easy.

A few studies have emphasized the need for organizations to look at strategies for dealing with external pressures related to the ethical dilemmas that can grow out of data collection (Günther *et al.*, 2017), such as privacy breaches. This study is an effort to fill the gap in the literature by offering an EA framework to ensure that the ethical and privacy features, as highlighted by Batistič and van der Laken (2019), are 'designed into a system before implementation begins' (Stallings, 2020: 316).

In this experimental study, we examined a broad range of websites where significant numbers of users register and supply personal information while setting up an account. We initially looked at the degree of leakage of private information via different sites, focusing on the direct leakage of private information (e.g. name, email address and gender) to third-party aggregators. Our study demonstrates that there are major concerns regarding security and privacy issues associated with online activities in the era of big data and cloud computing. The results show that multi-link web services, offered by many organizations, may result in a leak of users' sensitive private information to the outside world.

Theoretical implications

There are many studies using the lenses of the resource-based view and dynamic capabilities to discuss the value of big data analytics in helping organizations develop a competitive advantage (Akter *et al.*, 2016; Gupta and George, 2016; Srinivasan and Swink, 2018). What is less understood is that organizations also need to look at strategies for dealing with the external pressure that comes with ethical dilemmas (Günther *et al.*, 2017). The present study is among the few to utilize a theory-based focus to address the challenges of privacy in big data research through experimental research.

With the rise of big data in the past decade, the number and scale of privacy issues have increased, largely due to a lack of understanding and functional testing of the capabilities of new developments in technology. As a result, there is a need for ways to integrate privacy into the implementation of security. This can be done by adopting the principles of the PbD framework based in EA.

The PbD framework is designed to address the gaps in existing privacy frameworks and respond to dramatic changes in the ICT landscape – in particular, emergent technologies such as big data, AI, the IoT and decentralized networks. This study is an effort to address the gap in existing EAs by incorporating the PbD framework into EA design and implementation. We believe that the TOGAF framework in particular, as a flexible and open source architecture, has the capacity to incorporate PbD as part of its architectural deployment method. As well, it is assumed by many that EA security features incorporate all privacy components, which does not lead to an understanding of privacy as its own domain. These factors allow us to understand the lack of privacy protection within EA frameworks and implementation methodologies, and therefore to provide a solution to integrate privacy within EA. The key theoretical contribution of this research is the application of PbD principles into concrete design and implementation, showing the framework's flexibility along with its alignment with the resource-based view of ethics and privacy. We argue that organizations need to consider privacy concerns, particularly in the era of big data and information sharing. Using the theoretical lens of the resource-based view, this research develops an EA to address the managerial challenges of privacy management.

Practical implications

Effective management of a firm's resources, including big data, is of growing importance. Organizations should take proactive approaches to protect users' private information, rather than reactive ones. This can be partially accomplished through a redesign of online user interfaces, as shown in our EA-based PbD framework. Unfortunately, existing EA models lack the important component of privacy. This dearth of privacy integration and awareness has led to the adoption of industry-standard frameworks that significantly fail to consider privacy as a necessity in its own right. For example, the Zachman framework does not define privacy within the six levels of concerns it outlines; nor does the TOGAF. Privacy requirements are as important as business and system requirements, and thus developing policies that limit access based on the login method is valuable. Embedding privacy measures into the design would allow for the building of privacy directly into ICT, and applying more privacy to an application when logging in with a secure partner (as we suggested in the Biz case study above) will reduce privacy issues. It is beneficial for websites to add more security features and embed privacy issues into their code.

Limitations and conclusion

Our proposed privacy control and management model was limited to analysis of data extracted from websites; this being the case, other sources of data – such as cloud-based applications (e.g. Dropbox), sensor data from the IoT and mobile application data – were not included in this study. It is important that future studies expand the project to cover the above big data sources.

Big data analytics have been extensively studied as a way for organizations to gain market insights and remain competitive. What is less well researched is how organizations should mitigate the ethical concerns big data brings. Using an experimental research design, the current study aims to fill the gap in existing privacy frameworks and respond to dramatic changes in the ICT landscape – in particular emergent technologies such as big data, AI, the IoT and decentralized networks – by means of a PbD framework within the context of EA. Our theory-focused research aims to address the challenge of privacy in big data research through an experimental study of the most important privacy challenges posed by big data. The findings of the present research suggest a proactive approach to protecting users' private information, which can be partially accomplished through redesigning online user interfaces to incorporate our EA-based PbD framework.

References

Agre, P. E. and M. Rotenberg (Eds.). (1998). *Technology and privacy: The new landscape*. Mit Press.

Aier, S. (2014). 'The role of organizational culture for grounding, management, guidance and effectiveness of enterprise architecture principles', *Information Systems and e-Business Management*, **12**, pp. 43–70.

Akter, S., S. F. Wamba, A. Gunasekaran, R. Dubey and S. J. Childe (2016). 'How to improve firm performance using big data analytics capability and business strategy alignment?', *International Journal of Production Economics*, **182**, pp. 113–131.

Aydiner, A. S., E. Tatoglu, E. Bayraktar, S. Zaim and D. Delen (2019). 'Business analytics and firm performance: the mediating role of business process performance', *Journal of Business Research*, **96**, pp. 228–237.

Barney, J. (1991). 'Firm resources and sustained competitive advantage', *Journal of Management*, **17**, pp. 99–120.

Batistič, S. and P. van der Laken (2019). 'History, evolution and future of big data and analytics: a bibliometric analysis of its relationship to performance in organizations', *British Journal of Management*, **30**, pp. 229–251.

Bell, E. and A. Bryman (2007). 'The ethics of management research: an exploratory content analysis', *British Journal of Management*, **18**, pp. 63–77.

Bennett, C. J. (2011). 'In defence of privacy: the concept and the regime', *Surveillance and Society*, **8**, pp. 485–496.

Bennett, C. J. and C. D. Raab (2006). *The Governance of Privacy: Policy Instruments in Global Perspective*. Routledge.

Boyd, D. and K. Crawford (2012). 'Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon', *Information, Communication & Society*, **15**, pp. 662–679.

Carona, X., R. Bosuab, S. Maynardb and A. Ahmadb (2016). 'The Internet of Things (IoT) and its impact on individual privacy: an Australian perspective', *Computer Law & Security Review*, **32**, pp. 4–15.

Cavoukian, A. (2009). Privacy by design. Take the challenge. Information and privacy commissioner of Ontario, Canada.

Chamorro-Premuzic, T., F. Polli and B. Dattner (2019). 'Building ethical AI for talent management', *Harvard Business Review*. Available at: https://hbr.org/2019/11/building-ethical-ai-for-talent-management

Chari, S. (2014). High Value Insights with Big Data Analytics on IBM Power Systems. Available at: http://www03.ibm.com/systems/data/flash/ae/migratetoibm/res/High_Value_Insights_with_Big_Data_Analytics_on_Power.pdf (Accessed November 15, 2018).

Chen, D. Q., D. S. Preston and M. Swink (2015). 'How the use of big data analytics affects value creation in supply chain management', *Journal of Management Information Systems*, **32**, pp. 4–39.

Chen, H., R. Chiang and V. Storey (2012). 'Business intelligence and analytics: from big data to big impact', *MIS Quarterly*, **36**, pp. 1165–1188.

Clarke, R. (1997). *Introduction to Dataveillance and Information Privacy, and Definitions of Terms*. Available at: http://www.rogerclarke.com/DV/Intro.html (Accessed June 12, 2019).

Clarke, R. (2000). *Beyond theOECD Guidelines: Privacy Protection for the 21st Century*. Available at: http://www.rogerclarke.com/DV/PP21C.html

Cohen, M. C. (2018). 'Big data and service operations', *Production and Operations Management*, **27**, pp. 1709–1723.

Constantiou, I. D. and J. Kallinikos (2015). 'New games, new rules: big data and the changing context of strategy', *Journal of Information Technology*, **30**, pp. 44–57.

Culnan, M. J. and C. C. Williams (2009). 'How ethics can enhance organizational privacy: lessons from the choicepoint and TJX data breaches', *MIS Quarterly*, **33**, pp. 673–687.

Culnan, M. J., E. R. Foxman and A. W. Ray (2008). 'Why IT executives should help employees secure their home computers', *MIS Quarterly Executive*, **7**, pp. 49–55.

Dam, K. H., S. L. Lê and A. Ghose (2016). 'Managing changes in the enterprise architecture modeling context', *Enterprise Information Systems*, **10**, pp. 666–696.

Dattner, B., T. Chamorro-Premuzic, R. Buchband and L. Schettler (2019). 'The legal and ethical implications of using AI in hiring', *Harvard Business Review*. Available at: https://hbr.org/2019/04/the-legal-and-ethical-implications-of-using-ai-in-hiring [accessed 26 January 2020].

De Cremer, D., R. Van Dick, A. Tenbrunsel, M. Pillutla and J. K. Murnighan (2011). 'Understanding ethical behavior and decision making in management: a behavioural business ethics approach', *British Journal of Management*, **22**, pp. S1–S4.

Dhillon, G., R. Syed and C. Pedron (2016). 'Interpreting information security culture: an organizational transformation case study', *Computers & Security*, **56**, pp. 63–69.

Dinev, T., H. Xu, J. H. Smith and P. Hart (2013). 'Information privacy and correlates: an empirical attempt to bridge and distinguish privacy-related concepts', *European Journal of Information Systems*, **22**, pp. 295–316.

Dubey, R., A. Gunasekaran, S. J. Childe, C. Blome and T. Papadopoulos (2019). 'Big data and predictive analytics and manufacturing performance: integrating institutional theory, resource-based view and big data culture', *British Journal of Management*, **30**, pp. 341–361.

EDPS (2018). *European Data Protection Supervisor: Preliminary Opinion on Privacy by Design*. Available at: https://edps.europa.eu/sites/edp/files/publication/18-05-31_preliminary_opinion_on_privacy_by_design_en_0.pdf (Accessed November 15, 2018).

Eisenhardt, K. M. and J. A. Martin (2000). 'Dynamic capabilities: what are they?', *Strategic Management Journal*, **21**, pp. 1105–1121.

Elahi, S., M. d'Aquin and E. Motta (2010). 'Who want a piece of me? Reconstructing a user profile from personal web activity logs'. In *International ESWC Workshop on Linking of User Profiles and Applications in the Social Semantic Web*.

Engelsman, W., D. Quartel, H. Jonkers and V. M. Sinderen (2011). 'Extending enterprise architecture modeling with business goals and requirements', *Enterprise Information Systems*, **5**, pp. 9–36.

Finn, R. L., D. Wright and M. Friedewald (2013). 'Seven types of privacy'. In S. Gutwirth, R. Leenes, P. de Hert and Y. Poullet (eds), *European Data Protection: Coming of Age*, pp. 3–32. Dordrecht: Springer.

Fischer, R., S. Aier and R. Winter (2007). 'A federated approach to enterprise architecture model maintenance', *Enterprise Modeling and Information Systems Architectures*, **2**, pp. 14–22.

Foorthuis, R., M. Van Steenbergen, S. Brinkkemper and W. A. Bruls (2016). 'A theory building study of enterprise architecture practices and benefits', *Information Systems Frontiers*, **18**, pp. 541–564.

Günther, W. A., M. H. R. Mehrizi, M. Huysman and F. Feldberg (2017). 'Debating big data: a literature review on realizing value from big data', *The Journal of Strategic Information Systems*, **26**, pp. 191–209.

Gupta, M. and J. F. George (2016). 'Toward the development of a big data analytics capability', *Information & Management*, **53**, pp. 1049–1064.

Herschel, R. and V. M. Miori (2017). 'Ethics & big data', *Technology in Society*, **49**, pp. 31–36.

Lankhorst, M. (2009). *Enterprise Architecture at Work*. Berlin: Springer.

Li, Y., Y. Li, Q. Yan and R. H. Deng (2015). 'Privacy leakage analysis in online social networks', *Computers & Security*, **49**, pp. 239–254.

Lindgreen, A. and V. Swaen (2009), 'Corporate social responsibility', *International Journal of Management Reviews*, **12**, pp. 1–7.

Lindgreen, A., V. Swaen and W. J. Johnston (2009). 'Corporate social responsibility: an empirical investigation of U.S. organizations', *Journal of Business Ethics*, **85**, pp. 303–323.

Loch, C. H. and Y. Wu (2008). 'Social preferences and supply chain performance: an experimental study', *Management Science*, **54**, pp. 1835–1849.

McGovern, J., S. W. Ambler, M. E. Stevens, J. Linn, E. K. Jo and V. Sharan (2004). *A Practical Guide to Enterprise Architecture*. Prentice Hall Professional.

McKelvie, A. and P. Davidsson (2009). 'From resource base to dynamic capabilities: an investigation of new firms', *British Journal of Management*, **20**, pp. S63–S80.

Mikalef, P., M. Boura, G. Lekakos and J. Krogstie (2019). 'Big data analytics capabilities and innovation: the mediating role of dynamic capabilities and moderating effect of the environment', *British Journal of Management*, **30**, pp. 272–298.

Nikpay, F., R. B. Ahmad, B. D. Rouhani, M. N. R. Mahrin and S. Shamshirband (2017). 'An effective enterprise architecture implementation methodology', *Information Systems and e-Business Management*, **15**, pp. 927–962.

NIST (2010). National Institute of Standards and Technology Special Publication 800-122: Guide to Protecting the Confidentiality of Personally Identifiable Information (PII). Available at: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-122.pdf (Accessed June 20, 2019).

Nunan, D. and M. Di Domenico (2013). 'Market research and the ethics of big data', *International Journal of Market Research*, **55**, pp. 505–520.

Pinkston, T. S. and A. B. Carroll (1994). 'Corporate citizenship perspectives and foreign direct investment in the U.S.', *Journal of Business Ethics*, **13**, pp. 157–169.

Proper, E. and D. Greefhorst (2010). 'The roles of principles in enterprise architecture'. In M. Lankhorst, M. Schönherr, J. Barjis and S. Overbeek (eds), *Trends in Enterprise Architecture Research*, pp. 57–70. Berlin: Springer.

O'Leary, D. E. (2016). 'Ethics for big data and analytics', *IEEE Intelligent Systems*, **31**, pp. 81–84.

Owen, M. and J. Raj (2003). '*BPMN and Business Process Management: Introduction to the New Business Process Modeling Standard*'. Available at: http://www.bptrends.com/publicationfiles/03-04_WP_BPMN_and_BPM_Owen-Raj.pdf (Accessed July 2, 2019).

Quartel, D., W. Engelsman, H. Jonkers and V. M. Sinderen (2009, September). A goal-oriented requirements modeling language for enterprise architecture. In *2009 IEEE International Enterprise Distributed Object Computing Conference* (pp. 3–13). IEEE.

Ross, J. W., P. Weill and D. Robertson (2006). *Enterprise Architecture as Strategy: Creating a Foundation for Business Execution*. Harvard Business Press.

Rouhani, B. D., M. N. Mahrin, F. Nikpay and P. Nikfard (2013). 'A comparison of enterprise architecture implementation methodologies'. In *2013 International Conference on Informatics and Creative Multimedia*.

Rouhani, B. D., M. N. R. Mahrin, F. Nikpay, M. K. Najafabadi and P. Nikfard (2015). 'A framework for evaluation of enterprise architecture implementation methodologies', *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, **9**, pp. 1–6.

Sessions, R. (2007). 'A comparison of the top four enterprise-architecture methodologies'. Microsoft Developer Network Architecture Center. Available at: (Accessed June 24, 2019).

Shah, H. and M. El Kourdi (2007). 'Frameworks for enterprise architecture', *IT Professional*, **9**, pp. 36–41.

Solove, D. J. (2008). *Understanding Privacy*. Harvard University Press.

Srinivasan, R. and M. Swink (2018). 'An investigation of visibility and flexibility as complements to supply chain analytics: an organizational information processing theory perspective', *Production and Operations Management*, **27**, pp. 1849–1867.

Stallings, W. (2020). '*Information Privacy Engineering and Privacy by Design*'. Addison-Wesley.

Tamm, T., P. Seddon, G. Shanks and P. Reynolds (2011). 'How does enterprise architecture add value to organisations?', *Communications of the Association for Information Systems*, **28**, Article 10.

Tavani, H. T. and J. H. Moor (2001). 'Privacy protection, control of information, and privacy-enhancing technologies', *ACM SIGCAS Computers and Society*, **31**, pp. 6–11.

Teece, D. J., G. Pisano and A. Shuen (1997). 'Dynamic capabilities and strategic management', *Strategic Management Journal*, **18**, pp. 509–533.

The Buzz (2017, September 14). 'Equifax shares plunge again – 35% in past week'. Available at: https://money.cnn.com/2017/09/14/investing/equifax-stock/index.html

The Guardian (2017, September 25). 'Deloitte hit by cyber-attack revealing clients' secret emails'. Available at: https://www.theguardian.com/business/2017/sep/25/deloitte-hit-by-cyber-attack-revealing-clients-secret-emails

The New York Times (2018, September 28). ‘Facebook security breach exposes accounts of 50 million users’. Available at: https://www.nytimes.com/2018/09/28/technology/facebook-hack-data-breach.html

The Washington Post (2017, August 19). ‘How to search the Ashley Madison leak’. Available at: https://www.washingtonpost.com/news/the-intersect/wp/2015/08/19/how-to-see-if-you-or-your-spouse-appear-in-the-ashley-madison-leak/

Treviño, L. K., G. R. Weaver and S. J. Reynolds (2006). ‘Behavioral ethics in organizations: a review’, *Journal of Management*, **32**, pp. 951–990.

Tsai, J. Y., S. Egelman, L. Cranor and A. Acquisti (2011). ‘The effect of online privacy information on purchasing behavior: an experimental study’, *Information Systems Research*, **22**, pp. 254–268.

Ullah, S., S. Ahmad, S. Akbar and D. Kodwani (2019). ‘International evidence on the determinants of organizational ethical vulnerability’, *British Journal of Management*, **30**, pp. 668–691.

Venkatraman, N. and A. Zaheer (1990). ‘Electronic integration and strategic advantage: a quasi-experimental study in the insurance industry’, *Information Systems Research*, **1**, pp. 377–393.

Wamba, S. F., S. Akter, A. Edwards, G. Chopin and D. Gnanzou (2015). ‘How “big data” can make big impact: findings from a systematic review and a longitudinal case study’, *International Journal of Production Economics*, **165**, pp. 234–246.

Wired (2015, October11). ‘Four indicted in massive JP Morgan Chase hack’. Available at: https://www.wired.com/2015/11/four-indicted-in-massive-jp-morgan-chase-hack/

Wright, D. and C. Raab (2014). ‘Privacy principles, risks and harms’, *International Review of Law, Computers & Technology*, **28**, pp. 277–298.

Xu, L., C. Jiang, N. He, Z. Han and A. Benslimane (2019). ‘Trust-based collaborative privacy management in online social networks’, *IEEE Transactions on Information Forensics and Security*, **14**, pp. 48–60.

Yu, E., M. Strohmaier and X. Deng (2006). *Exploring Intentional Modeling and Analysis for Enterprise Architecture*. Available at: https://doi.org/10.1109/EDOCW.2006.36 (Accessed July 17, 2019).

Zachman, J. (2008). *John Zachman's Concise Definition of the Zachman Framework*. Available at: https ://www.zachman.com/about-the-zachman-framework (Accessed July 17, 2019).

Zahra, S. A., H. J. Sapienza and P. Davidsson (2006). ‘Entrepreneurship and dynamic capabilities: a review, model and research agenda’, *Journal of Management Studies*, **43**, pp. 917–955.

Zhang, O. Q., R. K. Ko, M. Kirchberg, C. H. Suen, P. Jagadpramana and B. S. Lee (2012). ‘How to track your data: rule-based data provenance tracing algorithms’. In *Proceedings of 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications*.

Nick Hajli is an Associate Professor at Swansea University. He has a PhD in Management from Birkbeck, University of London. His research has appeared in the top 20 journals used in Business School Research Rankings. Nick’s research interests include big data, social commerce, artificial intelligence and ethical issues surrounding emerging technologies. He has published over 80 peer-reviewed research papers in international journals. Nick also sits on the editorial board of several academic journals.

Farid Shirazi is a Senior Researcher at the Institute for Innovation and Technology Management at Ryerson University, Toronto, Canada. He is an Associate Professor of the Ted Rogers School of Information Technology Management. Farid’s research focuses mainly on the impact of ICTs on social and economic development. His main research interests are IT-enabled sustainability and development, big data analytics, artificial intelligence and the ethical and security perspectives associated with the introduction and use of ICTs.

Mina Tajvidi is a Lecturer in Digital Marketing at Newcastle University, UK. Her research has appeared in the top 20 journals used in Business School Research Rankings. Mina’s research interests are in the areas of social media, digital marketing, branding and big data analytics. She also studies the ethical issues surrounding new technologies such as privacy, trust and information security.

Nurul Huda is a Professor at the School of Engineering Technology and Applied Science of Centennial College. He also teaches a Business Technology course at Ryerson University. He has published over 40 peer-reviewed research articles in international journals and conference proceedings. Nurul’s research interests include privacy-enhancing technologies, cloud computing with a focus on data-centre services, cybersecurity, IoT, architectures for mobile ad hoc networks and delay-tolerant networks.