Addae, Joyce Hoese and Brown, Michael and Sun, Xu and Towey, Dave and Radenkovic, Milena (2017) Measuring attitude towards personal data for adaptive cybersecurity. Information and Computer Security . ISSN 2056-4961

MEASURING ATTITUDE TOWARDS PERSONAL DATA FOR ADAPTIVE CYBERSECURITY

Authors
[Redacted for blind review]

Abstract

Purpose – *This paper presents an initial development of a Personal Data Attitude (PDA) measurement instrument based on established psychometric principles. The aim of the research was to develop a reliable measurement scale for quantifying and comparing attitudes towards personal data that can be incorporated into cybersecurity behavioral research models. Such a scale has become necessary for understanding individuals' attitudes towards specific sets of data as more technologies are being designed to harvest, collate, share and analyze personal data.*

Design/methodology/approach *– An initial set of 34 five-point Likert style items were developed with 8 sub-scales and administered to participants online. The data collected were subjected to Exploratory and Confirmatory factor analysis and MANOVA. The results are consistent with multi-dimensionality of attitude theories and suggest the adopted methodology for the study is appropriate for future research with a more representative sample.*

Findings – *Factor analysis of 247 responses identified six constructs of individuals' attitude towards personal data: Protective Behavior, Privacy Concerns, Cost-Benefit, Awareness, Responsibility and Security. This paper illustrates how the PDA scale can be a useful guide for information security research and design by briefly discussing the factor structure of the PDA and related results.*

Originality/value *– This study addresses a genuine gap in the research by taking the first step towards establishing empirical evidence for dimensions underlying personal data attitudes. It also adds a significant benchmark to a growing body of literature on understanding and modelling computer users' security behaviors.*

Keywords: Personal data, Information disclosure, Privacy, Attitude scale, Cybersecurity, Information security modelling, Individual behavior

1. Introduction

The amount of personal data being captured, collated, analyzed and shared is growing every day. The trend is mediated by technologies such as smartphones, social networking, and smart-meters (Manyika et al., 2015). Previous work has highlighted that personal privacy is more vulnerable to erosions when contextual and personal information is gathered by pervasive computing systems (Beresford, 2005, Zhu et al., 2009). Although personal data is recognized as a key issue requiring innovative cybersecurity measures within the digital economy, there are comparatively few studies exploring individuals' attitudes towards it. Iachello and Hong (2007) reviewed privacy related literature within the context of HCI and identified the need for a deeper understanding of individuals' attitudes towards the phenomena as a major challenge. Lederer et al. (2004) also acknowledged the fact that differing aspects of privacy pose a challenge for the design of usable systems. Essentially, there is not enough information available to guide stakeholders, including new technology designers and policy makers, in dealing with or addressing issues related to personal data. Innovative mobile information service providers are, for instance, faced with the question of how different users will respond to personal and context-aware services. Cybersecurity designers especially need to understand differing aspects of personal data issues to be able to develop systems that can adequately support values that constitute acceptable social behavior.

The lack of empirically identified factors influencing individuals' digital security behaviour presents a major challenge to addressing the human component of cybersecurity. As technologies become more personalized and context-aware, there is a need to understand the interfaces and functionality required to accommodate individual differences in both use and attitude. The study presented in this paper is part of our ongoing work in information security behavioural research towards addressing the critical issue of leveraging knowledge about individual differences for the design of usable and adaptive cybersecurity. To identify relevant determinants of cybersecurity practices and predict individuals' security behavior, we have presented an improved cybersecurity research model (Redacted for blind review) that integrates Planned Motivation Theory (PMT) with TAM, and enables a wider variety of factors influencing cybersecurity behavior to be explored (see Figure I). Our approach considers the Personal Data Ecosystem (PDE) and how external factors such as users' prior experiences and demographic characteristics could shape an individual's beliefs regarding the benefits and consequences of certain security-related behavior. We have identified the development of a Personal Data Attitudes (PDA) measurement scale as key to understanding and responding to people's views and attitude towards the data set around which digital technologies are built. Against this background, we determined to explore the different dimensions of attitudes towards

personal data with the aim of developing a PDA scale, and verifying the dimensions of attitudes toward different types of personal data. Thus, the paper focuses specifically on the development of a quantitative PDA measurement scale for the capture and analysis of attitudes across groups, contexts and datasets. Such an instrument could then be adopted in further pragmatic research activities to substantiate propositions in this area to offer theoretically informed guidelines for addressing personal data issues. Essentially, the results presented in this paper support the inclusion of personal data as a measurable variable in behavioral models being amplified for adaptive cybersecurity.

Figure I: Proposed Research Model for the Investigation and Segmentation of Cybersecurity Behaviors

2. Personal Data in the Digital Era

Technological advancement has the potential to enhance peoples' lives in many ways by facilitating the generation and sharing of knowledge. However, whenever we interact directly or indirectly with technologies, we leave behind data trails and digital footprints which can be used to generate information about our lives and activities. A considerable and increasing amount of information is gathered about us which are processed, stored, explored, shared, commercialized, and potentially misused by both public and private individuals and/or organizations. This raises concerns around issues such as privacy, security and other digital asset rights (Joinson et al., 2006). Consequently, when personal data is gathered, such as by pervasive computing devices, it is important to consider the range of potential implications for the individuals concerned. Previous research projects have explored personal data, mostly focusing on exploring people's attitude towards a single or limited subset of existing and/or near future technologies rather than the personal data itself. For instance, Brown (2013) adopted vignette-based survey to explore the social implications of data gathered through "Internet of things" (IoT) in homes. They highlight a range of concerns about the technical systems but in some cases, it is unclear if these are due to the interface, data collection, display or data being collected. Although research approaches that explore users' attitudes to an identified technology are useful in assessing users' perceptions of a technology's usage of sensitive resources, it is also valuable to look at attitudes towards the data itself rather than the technologies through which it is created and accessed (Sharples et al., 2013).

2.1 Underlying Dimensions of Personal Data

A requisite preliminary stage in the creation of a validated measurement instrument is a consideration of the relevant construct dimensions. The relevant construct in this case, is individuals' attitudes towards personal data. In defining attitude to personal data, we adopt the general definition for attitude from the behavioral science literature and applied it to our context. Attitude to personal data here therefore refers to the behavioral tendency of an individual to negatively or positively evaluate the disclosure of a personal data. To successfully capture the factors influencing individuals' attitude to personal data, existing literature on the underlying dimensions of the concept was first reviewed. We started by considering available definitions of the object towards which attitude is being measured — Personal Data. In the UK's Data Protection Act 1998 (ICO, 2015), personal data is defined as:

"data which relate to a living individual who can be identified (a) from those data, or (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller, and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual."

Personal information is also defined in the Privacy Protection Act of Australia as:

"...information or an opinion about an identified individual, or an individual who is reasonably identifiable: (a) whether the information or opinion is true or not; and (b) whether the information or opinion is recorded in a material form or not" (Australian Government, 2014, p. Subsection 6(1)).

The key component of these definitions is identifiability which means that certain level of personal information can be handled without legal implications as long as they are anonymized (Millard and Hon, 2012). Consequently any information that can reasonably be directly or indirectly linked to an individual's identity qualifies as a personal identifiable information and requires careful handling (US Department of Labor, 2015). Information considered to be linkable to individuals includes medical, educational and financial records (McCallister, 2010). With this background, we decided to focus on measuring and comparing attitude towards 4 main types of personal data (Social Media, Personal Emails, Financial and Health records).

In the process of reviewing existing literature on personal data, we realized that, although the concept has now become a hot topic in the cybersecurity and privacy research community, much of the focus has been on developing technical and legal countermeasures. Studies exploring the individual behavioral elements of the concept are quite limited hence, information on the structure of personal data as a psychometric construct is very scanty. Majority of the literature exploring personal data attitude based their studies on health and/or medical records. Rindfleisch (1997) for instance proposed and explained three concepts underlying health care information protection concerns – security, confidentiality and privacy. Wellcome Trust (2013) also measured and classified peoples' attitude to health data into identity, attention and control concerns. Consequently, three major themes were initially identified in the literature as dimensions of public views to personal data. These include issues related to security, risk/benefit trade-of and privacy/confidentiality. These dimensions mostly overlap with the eight underlying principles of the fair information practices outlined in the United States: PPSC (1977) and the (OECD, 2002) reports. The attention concern described by (WT, 2013) where users express mixed blessings about giving away their personal information, falls within the object of purpose specification principle. Thus, a user may give out personal information for an immediate benefit but may not be sure of the future implications of such an action. The principle of use limitation extends to the case where a user may be concerned about their identity being abused once their personal data is disclosed to access a service. The principle of collection limitation aims to addresses security and privacy concerns, by providing a framework for limiting the amount of specific PII that can be collected within specific contexts. Most definitions of privacy refers to people's ability to control the terms under which their personal information are acquired and used (e.g. Westin, 1967, Rindfleisch, 1997, Earp and Payton, 2006). Collection limitation is about having a limit to the collection of personal data and doing so in a legalized manner. The dimension of control is illustrated as a personal data concern with users' comments on losing their freewill and not being able to go for 'opt-outs' as illustrated in the Wellcome Trust report.

The World Economic Forum (WEF, 2013, p. 3) hosted a global dialogue on the emerging issues surrounding the collection and use of personal data by clustering the eight principles into three main themes: *"Protection and security, Accountability and Rights and responsibilities for using personal data"* Security here has to do with the integrity, availability and controlled access to information. This clearly encompasses the control concern identified by Wellcome Trust. The concept of confidentiality described by (Rindfleisch, 1997) directly overlaps with the principle of use limitations as they both has to do with the release of information when accessing a service (in this case health care) in a legal manner that limits the extent to which they may further be used or released. Rindfleisch (1997) refers to privacy as the right and desire of a person to control the disclosure of personal health information. This description is very much synonymous to the general definitions of privacy identified in the literature as mentioned earlier. Though there are limited studies attempting to measure attitude to personal data directly, several studies have explored privacy concerns. Smith et al. (1996) developed the Concern for Information Privacy (CFIP) scale which identified and measured four factors (collection, errors, secondary use and unauthorized access to information) as the dimensions of a person's privacy concern about organizations. Malhotra et al. (2004) later identified three aspects of information privacy namely: attitudes towards the collection of personal information, control over personal information and awareness of privacy practices. More recently Hong and Thong (2013), consolidated existing conceptualization of Internet Privacy Concerns (IPC) including (Smith et al., 1996) and (Malhotra et al., 2004) in their study and came up with six first-order factors of IPC – collection, secondary usage, errors, improper access, control, and awareness. When reviewing these IPC measurement instruments, this article focuses on the broader personal data ecosystem by identifying constructs that goes beyond privacy concerns to encompass other construct domains. Table ITable I summarizes the construct domains generated from the review of existing literature pertaining to attitudes toward personal data.

Table I: Potential Constructs of Attitude to Personal Data

3. Scale Development

This paper describes a multistage scale development study conducted to develop a measurement instrument for PDA. Fundamentally, the issues of reliability and validity underpin the development of an attitude scale right from item generation, the theoretical deduction of a factor structure and resulting psychometric analysis. Previous scale development studies as well as recommendations from (Hinkin, 1995) on how to improve the scale development process provided guidance for the research. Subsequently, an iterative process (see Figure II) is adopted to assess the consistency of the scale

items with the dimensions of personal data attitudes identified in the literature. This section describes the procedures involved in developing and assessing the initial version of the PDA instrument.

Figure II: Iterative work-flow adopted for the development of a PDA measurement scale

3.1 Generation of Initial Pool of PDA Items

Following an examination of existing personal data related literature, definitions and surveys (including Joinson et al., 2006, Lang et al., 2009), 50 Likert-type attitudinal items related to the 10 construct domains identified was generated. The items drawn from these sources were rephrased to mainly reflect attitudes towards four different types of personal data namely: Email, Social Media (online personal data); Financial and Health (offline personal records). Since we did not find a measurement scale specifically designed for attitude towards personal data in the existing literature, an exploratory study was conducted as a preliminary step toward generating the PDA scale items. To ensure the content validity of construct domains predetermined from the literature review, focus group discussions were held using open-ended questions to elicit themes that constituted individuals' view on the four types of personal data. 50 additional items were generated based on responses emerging from the focus group discussion on personal data matters. This resulted in an initial pool of 100 items serving as the basis for the PDA measurement. 64 items were eventually dropped following an exercise to merge similar themes and convert the 100 items into generic personal data statements. Thus, those that were obviously pointing to a specific type of personal data (e.g. I feel my profile information on any social media is much secured) were discarded.

3.2 Scale Specification and Refinement

The research team reviewed the remaining 46 items based on the 10 construct definitions in Table I~~Table I~~. We adapted items from previous privacy related measurement scales to fit the PDA context as much as it was possible to do so. Items generated for the Experience and Sensitivity construct domains were dropped as they were mainly measured with categorical rather than scale data in previous studies. To minimize the tendency for respondents to provide the same responses or agree with statement due to acquiescent response bias, some items were worded negatively. To do this the direction of each statement (positive or negative attitude to personal data) needed to be determined. Items that were categorized as unable to judge statements were discarded. For instance, we could not indicate whether a statement like '*I do not mind sharing such information with family and friends'* is a positive or negative attitude towards personal data. Consequently, the items were further reduced to 34 PDA statements.

Each of the items was a statement to which people were asked about their level of agreement on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). Thus, attitude to personal data was quantified as a continuous variable. Questions relating to some of the theoretically distinct aspects of personal data discussed earlier were included. For instance, items relating to security/identity (e.g. '*There should be stronger laws to protect such personal data'*), control/privacy (e.g. *'I consider the privacy policy of institutions where I give out such personal details'*) and possible benefits of surrendering personal data (e.g. *'I am happy to provide such personal details to support government policy and decision making'*). The scale involved both positively and negatively worded items. To ensure that a higher numbered response on the Likert scale would represent positive attitudes, all negatively worded items were reversed before the data was analyzed.

3.3 Data Collection

A web-based questionnaire was developed based on the 34 PDA items. The questionnaire had two main sections of demographics and the attitude to personal data items with four research design conditions (Personal Email, Social Media, Financial and Health data). Essentially participants were randomly presented the generic statements with respect to one of the four types of personal data outlined above until they were almost evenly distributed across the groups. A non-probability sampling approach was adopted to collect responses from participants over the internet. Email invitations and anonymous link to the survey was posted on social media sites (Facebook, Twitter and LinkedIn). To further expand the sample size, snow-balling and convenience sampling techniques were also used. After several follow-up rounds, a total of 256 responses was received out of which 247 completed datasets was extracted at the data cleaning and preparation stage. Of the 247 respondents, 51.4% (127) were male and 48.6% (120) were female. The average age of the sample was 36 years (range: 17 – 67 years). As presented in Table II, most respondents use Social Media (90.7%) out of which most of them preferred to use Facebook (47.8%). Respondents who have not had previous experience with personal information misuse formed a significant portion of the sample population (83%).

Table II: Participant demographics

4. Analysis and Results

4.1 Reliability Analysis

The 34 items of the attitude to personal data measure were subjected to an iterative scale purification procedure. To determine the internal consistency of the items, the most widely used reliability method of computing the Cronbach's alpha was adopted (Papanastasiou, 2005). This yielded 0.926, indicating a high reliability of the 34-item scale. A correlation matrix generated with SPSS was scanned to check the pattern of relationships among the items. There is no singularity in the data as all the correlation coefficients were less than 0.8 (Field 2013). Another commonly accepted procedure used to further assess the internal consistency of the items was the item-total correlations. A close look at the inter-item correlations and item-to-total correlations of each item revealed inadequate performance of some items. If the items are all measuring attitude to personal data, then each item ought to correlate with the total score from the questionnaire (Rattray and Jones, 2007). The correlation between participants' score on an item and the sum of their scores on all the items is represented by the *r* value. Three (3) of the items that poorly correlated ($r<0.4$) were removed from the PDA items. This conforms to the generally accepted rule-of-thumb that item-to-total correlations should exceed 0.30 (Van der Heijden and Sørensen, 2002). After deleting the items that fell below this standard, 31 items remained in the pre-final version of the questionnaire. The 31 items were then subjected to a separate reliability test. This resulted in an acceptable item-total correlation but the Cronbach alpha remained at 0.926.

4.2 Exploratory Factor Analysis (EFA)

An exploratory factor analysis can help to empirically determine how many constructs, or factors, underlie the set of PDA items (Child, 2006, Rattray and Jones, 2007). To determine the appropriateness of the factor analysis, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's Test of Sphericity were first examined. The KMO value for the data set was 0.899 and Bartlett's Test of Sphericity was significant ($p < 0.000$), indicating that the factor analysis was appropriate. The initial EFA analysis produced a pattern matrix consisting of seven factors based on Eigenvalues greater than one and accounted for 64.135% of the total variance. Only one item loaded on the 7th factor and most of the items generated for Protective Behavior and the Interest Scale were loading together. To minimize errors associated with under-extraction and/or over-extraction, a mixed approach based on both the eigenvalue and scree plot results was adopted. For instance, under-extraction error could lead to inconsistencies in the analysis and interpretation of the results (Reise et al., 2000). In effect, based on the initial factor extraction and the results from the scree test (see Figure III), the factor analysis was repeated to extract a 6-factor solution with an oblique rotation method to allow the obtained components to correlate (see Table III). This supports the assumption that PDA dimensions are related yet distinct from each other (Lee and Comrey, 1979). The final six-factor model emerging from the 31 PDA items explained 63.983% of the total variance. Factor 1 contained the items measuring protective behaviour and interest, while factor 2 focused on the items measuring privacy. Factor 3 involved the items measuring cost-benefit and factor 4 contained the items measuring awareness. Factor 5 consist of the items measuring responsibility, and factor 6 involve those items measuring security.

Figure III: Scree plot of factors underlying the PDA scale.

4.3 Assessment of the Factor Structure

Next, Confirmatory Factor Analysis (CFA) was conducted using AMOS 23 to examine the factor structure obtained from the EFA. Maximum likelihood parameter estimates were used to examine the model fit. Researchers are required to report several fit indices in order to characterize the fitness of a model correctly with the following boundary scores indicating good model fit – $\chi^2/df = 2.0–5.0$, RMSEA< 0.06 [<0.05, < 0.08] and CFI >0.94 (Hooper et al., 2008). Results of the CFA ($\chi^2/df = 1.802$, RMSEA = 0.057 and CFI = 0.940) indicate that the measurement model fits the data quite well. The 6-factor solution is therefore supported by the CFA results. As observed in Table III, the reliability of the PDA scale is supported by the composite reliability estimates (> 0.7) ranging from 0.84 to 0.93, indicating a good internal consistency of the multiple items for each construct in the model. The convergent validity was evaluated by checking all the values of average variance extracted (AVE) and the factor loadings. As shown in Table III, the estimated AVEs of all the PDA

dimensions were all greater than the unexplained variances (>0.5) and all the factor loadings for the six constructs were above 0.5 and were significant for the individual items. The examination of the AVEs together with the factor loadings therefore confirmed the convergent validity of the PDA latent constructs. To investigate the discriminant validity of the PDA scale, the suggestions provided by Bertea and Zait (2011) regarding use of AVE analysis were followed. Accordingly, the value of the AVE for each construct should be at least 0.50 and the square root of each construct's AVE should be much larger than the correlation of the specific construct with any other constructs. The results from the AVE analysis presented in Table IV show that the shared variance between any two constructs was not greater than the square root of the corresponding AVEs. In summary, the assessments carried out to verify the measurement model yielded evidence for the reliability of the latent constructs.

Table III: Scale Items and CFA Results

Table IV: Factor Inter-correlations

4.4 Analysis of Variance

The data was also analyzed to determine whether there are differences between the four types of personal data. A series of Multivariate Analysis of Variance (MANOVA) was performed using factor scores for participants' responses on the six constructs identified above as the dependent variables (DVs); and type of personal data (Email, Health Data, Financial Records and Social Media), prior experience (Figure IV~~Figure IV~~) and sensitivity responses (Yes and No) as the independent variables (IVs). — see Figure V~~Figure V~~. Bartlett's approach was used to compute the factor scores in SPSS to obtain unbiased estimates of the true factor scores (Everitt and Howell, 2005). As shown in Table V, the multivariate test using WilksLambda revealed an overall significant effect for data type as an IV on the DVs ($F= 1.77$) at $p = 0.025$ ($p < 0.05$). Univariate analysis for the effect of type of personal data in the survey, presented in Table VI, significantly predicted responses related to **Behavior** ($p<0.05$) and **Privacy Concerns** ($p<0.05$). Finally, post hoc testing (with Least Significant Difference (LSD)) revealed highest **Protective Behavior** scores for Health Data (mean = 24.87, SD = 3.03) and Email (mean = 24.19, SD = 3.44), with Health Data being significantly higher (at $p<0.05$) than Social Media (mean = 23.51, SD = 3.72). No significant mean difference in Privacy Concerns scores were obtained between the four types of personal data but the highest score was for **Financial Records** (mean= 21.58, SD= 5.16). The implications of these findings are discussed in the next section.

Figure IV: A frequency distribution of responses per personal data type and prior misuse experience

Figure V: A frequency distribution of responses per personal data type and perception of sensitivity of data type

Table V: Multivariate ANOVA done on the factor scores for the 6 PDA constructs across 4 types of personal data

Table VI: Univariate Test done on each of the 6 DVs for the types of personal data

4.5 Cluster Analysis

To identify homogeneous groups in the dataset, we used a TwoStep clustering approach to cluster participants based on the six PDA factor scores computed. TwoStep Clustering was chosen due to its ability to automatically determine the optimal number of clusters in the dataset using an agglomerative hierarchal method (Mooi and Sarstedt, 2010). BIC (Schwarz's Bayesian Information Criterion) was used first to determine the number of clusters, then AIC (Akaike's Information Criterion) was used. Table VII summarizes the results obtained with BIC, which do not differ from those obtained with

AIC. The cluster centroids facilitated interpretation of the resultant cluster solution. The resulting clusters are labelled based on the Protective Behavior (PB) factor which happens to be the overall important predictor variable (see Figure VI). The first cluster, which is the largest (57.9%), contains participants with responses indicating a general privacy consciousness (mean Privacy factor score = 0.11) and a somewhat protective attitude towards their personal data, with a mean score of 0.06 on the PB factor. The second cluster contain 23.9% of the total participants with responses indicating high Awareness on Personal Data issues, and obtained the highest mean score on the PB factor (0.84). The third cluster is the smallest (18.2%) and consist mainly of participants whose responses indicate a general lack of interest in Personal Data related issues, and scored the lowest on all six PDA factors, especially on PB (Mean = -1.30). The cluster centers presented in Table VII are sorted to highlight the within-cluster predictor (PDA Factors) importance while Figure VI highlights the overall cluster membership predictor importance. A multivariate analysis conducted using the clusters as independent variable and the six factors as dependent variables shows significant differences in personal data attitudes across the segments (Wilks' lambda = 0.196, p>0.000). The results of a univariate F test revealed the clusters were significantly different on all PDA segment predictor factors. The discriminant analysis conducted based on the three clusters indicated that the model could correctly classify 94.3% of respondents into groups.

Table VII: Cluster distribution with cells showing cluster centers sorted by within-cluster membership predictor importance

Figure VI: Comparison of the relative distribution of PDA Factor Scores sorted by overall cluster membership predictor importance for the three clusters

5. Discussion

This study aimed to establish the dimensions required for the development of a reliable and valid PDA measurement scale. A PDA scale was successfully developed and verified based on the iterative scale development procedure adopted. Although the focus of this study was on attitude to personal data, the scale development process included an examination of both privacy and personal data literature, and the related regulations. The research results show that six constructs (Protective Behavior/Interest, Privacy, Cost-Benefit, Awareness, Security and Responsibility) are important components capable of differentiating individual attitude towards personal data. The reliability analysis yields support for the PDA scale's ability to reliably measure individual differences. The MANOVA results show that individuals' attitudes may vary based on the type of personal data. The results also suggest that, participants who provided responses to the scale items based on health records generally viewed this data set as sensitive, and tended to score higher on the protective behavior construct. Conversely, those who provided responses in relation to their personal social media profile data mostly did not view it as sensitive, and tended to score the lowest on both protective behavior and privacy concerns. Interestingly, there were a lot more social media participants who had had prior experience with personal data misuse (18) as compared to health data participants (4) – see Figure IV~~Figure IV~~.

The study also examined the relationship between the six extracted factors. Overall, the strongest relationship existed between the construct of Privacy concerns and Awareness as well as Privacy and Security. Thus, Privacy correlates positively and reliably with both attitudes relating to Awareness and Security concerns. A possible interpretation here is that people who are more concerned about their privacy being breached tend to be more aware of the potential value of their personal data, and the risk factors associated with it. In general, all the constructs correlate significantly with each other. Notwithstanding the assumption that each construct measured completely different aspects of individuals' personal data attitude, the level of correlation between them is an indication of the conflicting conceptions people generally have on the subject. The results from the cluster analysis show that, despite the diversity in attitudes towards personal data, there is a relatively small number of compatible groups of users sharing similar attitudes and behaviors. Research reported by Norberg et al. (2007) and (Sato, 2010) indicates that even though consumers express concerns for their personal data, they generally do not take enough protection measures towards it. Acquisti and Gross (2006) also compared stated attitudes with actual behaviors of members of online social networks (OSNs) and found that reported privacy attitudes did not correlate with the probability of disclosing certain type of information on OSNs. Perhaps exploration of the PDA constructs may assist researchers in examining the relationships between individuals' personal data concerns and their claimed personal interest and protective behaviors. For example, 91% of the participants in the study conducted by Sato (2010) indicated the need for a system that will enable them to control how their data is used. Meanwhile, researchers report fewer people actually adopting existing security and privacy mechanisms to protect their personal data online (EU Commission, 2011, Shelton et al., 2015). The research findings by Shelton et al. (2015) highlight usability and lack of awareness of existing

security mechanisms as the two main factors hindering people's ability to be more actively involved in the protection of their personal data and privacy. The usable security and privacy research community could therefore identify specific modifications of personal data attitudes, behaviors and skills that can foster a more positive appreciation for personal data through an in-depth exploration of the constructs identified.

The findings mostly reflect the underlying themes explored through studies focused on personal data, rather than those focused on privacy concerns. Thus, whereas privacy measurement scales tend to focus on constructs such as collection, control, errors, authorized use and awareness of privacy practices, the personal data literature is more concerned with issues relating to security, availability, privacy, risk and benefits. For instance, Kobsa (2007) suggest that although consumers appreciate the benefits of user profiling and personalization, they are not willing to be profiled due to privacy concerns. Sato (2010) on the other hand, surveyed about 3,000 people from six different countries and concluded that even though people are generally concerned about the privacy of their data, they are more positive about the potential benefits which they would normally weigh against the risk. Accordingly, when the benefits outweigh the perceived risk, cloud services users become more open to the data sharing concept. Although most people now accept that life in the digital age involves disclosure of personal data, concerns remain about the actual use of the data (EU Commission, 2011). However, as Acquisti et al. (2016) pointed out, because the experience individuals may have when their personal information is exposed may differ, their concerns about the use of their personal data also tend to vary. A PDA scale will therefore be required to capture these individual differences to enable more representative user-models for the design of cybersecurity tools.

The findings of the study have implications for both research and design practices in the field of usable security and privacy. Since little prior research exists specifically on scale development for PDA, this study signifies the first empirical examination of the concept. Existing instruments attempting to assess attitudes towards personal data are mainly based on privacy-focused design which have produced a variety of attitudes ranging from one to six dimensions. Although there are significant similarities in the attitudes between privacy and personal data, privacy instruments tend not to relate specifically to attitudes towards what people may view as personal information. Information privacy, one of the most vital aspects of privacy, is concerned with protecting personal data of individuals. However, the range of potential implications in relation to the collection and sharing of personal data goes beyond the issue of privacy and includes constructs related to responsibilities, security, risk and benefits as explored in this study. All stakeholders within the digital economy, need to carefully consider these dynamics to ensure that they understand and are willing to accept risk reward balance of personal data ecosystem (Kalapesi, 2013). Consequently, the study makes a major contribution to the growing body of literature on user modelling in the field of information security by highlighting the potential of including PDA as a determinant in predictive user models necessary for the design of adaptive cybersecurity.

In summary, the findings are believed to provide two major contributions to information security research and design practices: (I) a framework describing the primary dimensions of individuals' attitude towards personal data; and (II) an instrument that can easily be modified and used to measure those concerns and preferences for adaptive security designs. As indicated earlier, there is relatively little research literature on attitudes towards personal data that deal specifically with how individuals view the construct of personal data. There is therefore relatively little information available to guide designers in addressing personal data issues when designing new interfaces and technologies. A lot of new technologies and services have implications for how personal information is handled and how people react to them. Therefore, the existence of an instrument such as the PDA scale has the potential for distinctively detecting attitude profiles of technology users that can be very useful for adaptive cybersecurity designs. Essentially, the six factors identified provide a framework around which personal data discussions and models might be developed by human-computer interaction (HCI) researchers and information system designers. Thus, human factor engineers and designers may find such a tool very useful in looking to personalized interfaces where personal data issues are pertinent.

5.1 Study Limitations and Future Directions

Although the findings presented in this report form an effective first draft of a personal data attitude instrument, several limitations of the study need to be highlighted. Notably, the convenience and accidental sampling methods adopted may limit the external validity of the findings. This preliminary data has however been used to provide empirical evidence in support of the PDA scale's potential to be a valuable cybersecurity research and design tool. However, further research work need to be carried out with different types of populations, to establish the external validity of the PDA instruments. We recognize that our clustering approach is relatively basic and plan to extend the data collection towards refining the

clusters as part of our future work. Nevertheless, the primary contribution of this work is to demonstrate the feasibility of segmenting users based on their attitude towards personal data among other determinants.

The evaluation presented in this paper is the first step toward developing a robust empirical evidence of the PDA dimensions. Future research with more broader samples are required to replicate the factor structure and validate inferences that can be made based on the PDA scores. Variables like, technology users' personal data attitude change overtime and the underlying factors such as local context and/or cultural differences, could then be examined within a single integrated analysis using machine learning techniques and structural equation modelling. Finally, the findings of this study could be incorporated into the design of user models required to personalize cybersecurity or privacy education and products to different category of internet users.

Acknowledgments

[Redacted for blind review]

References

Acquisti, A. & Gross, R. Imagined communities: Awareness, information sharing, and privacy on the Facebook. International workshop on privacy enhancing technologies, 2006: Springer. 36-58.

Acquisti, A., Taylor, C. R. & Wagman, L. 2016. The economics of privacy. *Available at SSRN 2580411*.

Australian Government. 2014. *Privacy amendment (enhancing privacy protection) act 2012* [Online]. Australian Government. [Accessed].

Awad, N. F. & Krishnan, M. S. 2006. The personalization privacy paradox: an empirical evaluation of information transparency and the willingness to be profiled online for personalization. *MIS quarterly*, 13-28.

Bansal, G. & Gefen, D. 2010. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. *Decision Support Systems,* 49, 138-150.

Belanger, F., Hiller, J. S. & Smith, W. J. 2002. Trustworthiness in electronic commerce: the role of privacy, security, and site attributes. *The journal of strategic Information Systems,* 11, 245-270.

Beresford, A. R. 2005. Location privacy in ubiquitous computing. University of Cambridge, Computer Laboratory.

Bertea, P. & Zait, A. 2011. Methods for testing discriminant validity. *Management & Marketing-Craiova*, 217-224.

Brown, M. C. 2013. Exploring Interpretations of Data from the Internet of Things in the Home. *Interacting with Computers,* 3, 25.

Buchanan, T., Paine, C., Joinson, A. N. & Reips, U. D. 2007. Development of measures of online privacy concern and protection for use on the Internet. *Journal of the American Society for Information Science and Technology,* 58, 157-165.

Chen, L. F. & Ismail, R. 2013. Information Technology program students' awareness and perceptions towards personal data protection and privacy. *2013 International Conference on Research and Innovation in Information Systems (ICRIIS).*

Child, D. 2006. *The essentials of factor analysis,* London, Continuum.

Dinev, T. & Hart, P. 2006. An extended privacy calculus model for e-commerce transactions. *Information Systems Research,* 17, 61-80.

Earp, J. B. & Payton, F. C. 2006. Information privacy in the service sector: An exploratory study of health care and banking professionals. *Journal of Organizational Computing and Electronic Commerce,* 16, 105-122.

Eu Commission 2011. Attitudes on data protection and electronic identity in the European Union. *Eurobarometer Special Surveys,* 359.

Everitt, B. S. & Howell, D. C. 2005. *Encyclopedia of statistics in behavioral science*, John Wiley & Sons Ltd.

Hinkin, T. R. 1995. A review of scale development practices in the study of organizations. *Journal of management,* 21, 967-988.

Hon, W. K., Millard, C. & Walden, I. 2012. Who is responsible for 'personal data'in cloud computing?—The cloud of unknowing, Part 2. *International Data Privacy Law,* 2, 3-18.

Hong, W. & Thong, J. Y. 2013. Internet privacy concerns: an integrated conceptualization and four empirical studies. *MIS Quarterly,* 37, 275-298.

Hooper, D., Coughlan, J. & Mullen, M. 2008. Structural equation modelling: Guidelines for determining model fit. *Articles*, 2.

Hui, K.-L., Teo, H. H. & Lee, S.-Y. T. 2007. The value of privacy assurance: an exploratory field experiment. *Mis Quarterly*, 19-33.

Iachello, G. & Hong, J. 2007. End-user privacy in human-computer interaction. *Foundations and Trends in Human-Computer Interaction,* 1, 1-137.

Information Commissioner's Office 2015. Key definitions of the Data Protection Act. Information Commissioner's Office. ICO web page.

Joinson, A., Paine, C., Buchanan, T. & Reips, U. 2006. Measuring Internet Privacy Attitudes and Behavior: A multi-dimensional approach. *Journal of Information Science,* 32, 334-343.

Kalapesi, C. 2013. Unlocking the value of personal data: From collection to usage. *World Economic Forum technical report.*

Kobsa, A. 2007. Privacy-enhanced web personalization. *The adaptive web.* Springer.

Lang, M., Devitt, J., Kelly, S., Kinneen, A., O'malley, J. & Prunty, D. 2009. Social networking and personal data security: a study of attitudes and public awareness in Ireland. *In Management of e-Commerce and e-Government, ICMECG'09*.

Lederer, S., Hong, J. I., Dey, A. K. & Landay, J. A. 2004. Personal privacy through understanding and action: five pitfalls for designers. *Personal and Ubiquitous Computing,* 8**,** 440-454.

Lee, H. B. & Comrey, A. L. 1979. Distortions in a commonly used factor analytic procedure. *Multivariate Behavioral Research,* 14**,** 301-321.

Malhotra, N. K., Kim, S. S. & Agarwal, J. 2004. Internet users' information privacy concerns (IUIPC): The construct, the scale and a causal model. *Information Systems Research,* 5**,** 336-355.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. & Byers, A. H. 2015. Big data: The next frontier for innovation, competition, and productivity. 2011. McKinsey Global Institute

Mccallister, E. 2010. *Guide to protecting the confidentiality of personally identifiable information*, Diane Publishing.

Millard, C. & Hon, W. K. 2012. Defining 'personal data'in e-social science. *Information, Communication & Society,* 15**,** 66-84.

Mooi, E. & Sarstedt, M. 2010. *Cluster analysis*, Springer.

Norberg, P. A., Horne, D. R. & Horne, D. A. 2007. The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs,* 41**,** 100-126.

Opm 2015. Review of public and professional attitudes towards confidentiality of healthcare data. General Medical Council.

Organisation for Economic Co-Operation and Development 2002. *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*, OECD.

Papanastasiou, E. C. 2005. Factor structure of the "Attitudes Toward Research" scale. *Statistics Education Research Journal,* 4**,** 16-26.

Pattinson, M., Parsons, K., Butavicius, M., Mccormac, A., Calic, D., Furnell, S. & Furnell, S. 2016. Assessing Information Security Attitudes: A comparison of two studies. *Information & Computer Security,* 24.

Pearson, S. 2013. Privacy, security and trust in cloud computing. *Privacy and Security for Cloud Computing.* Springer.

Rattray, J. C. & Jones, M. C. 2007. Essential elements of questionnaire design and development. *Journal of Clinical Nursing,* 16**,** 234-243.

Reise, S. P., Waller, N. G. & Comrey, A. L. 2000. Factor analysis and scale revision. *Psychological Assessment,* 12**,** 287.

Rindfleisch, T. C. 1997. Privacy, information technology, and health care. *Communications of the ACM,* 40**,** 92-100.

Robling, M., Hood, K., Houston, H., Pill, R., Fay, J. & Evans, H. 2004. Public attitudes towards the use of primary care patient record data in medical research without consent: a qualitative study. *Journal of Medical Ethics,* 30**,** 104-109.

Sato, M. 2010. Personal data in the cloud: A global survey of consumer attitudes. Fujitsu Research Institute.

Sharples, S., Brown, M., Harding, J. & Jackson, M. 2013. Usability, human factors and geographic information (editorial). *Applied ergonomics,* 44**,** 853-854.

Shelton, M., Rainie, L. & Maddenn, M. 2015. Americans' Privacy Strategies Post-Snowden. Pew Research Center's Internet & American Life Project.

Sheng, H., Nah, F. F.-H. & Siau, K. 2008. An experimental study on ubiquitous commerce adoption: Impact of personalization and privacy concerns. *Journal of the Association for Information Systems,* 9**,** 344.

Singleton, P., Lea, N., Tapuria, A. & Kalra, D. 2008. Public and Professional attitudes to privacy of healthcare data: a survey of the literature. Cambridge Health Informatics Ltd

Smith, J. H., Milberg, S. J. & Burke, S. J. 1996. Information privacy: Measuring individuals concerns about organizational practices. *Management Information Systems Quarterly,* 20**,** 167-196.

United States: Ppsc 1977. *Personal privacy in an information society: the report of the Privacy Protection Study Commission*, The Commission: for sale by the Supt. of Docs., US Govt. Print. Off.

Us Department of Labor. 2015. *Guidance on the protection of personal identifiable information* [Online]. United States Department of Labor. [Accessed].

Van Der Heijden, H. & Sørensen, L. S. 2002. *Measuring attitudes towards mobile information services: an empirical validation of the HED/UT scale*, Technical University of Denmark, Center for Tele-Information.

Wang, Z. & Yu, Q. 2015. Privacy trust crisis of personal data in China in the era of Big Data: The survey and countermeasures. *Computer Law & Security Review,* 31**,** 782-792.

Wellcome Trust 2013. Summary Report of Qualitative Research Into Public Attitudes to Personal Data and Linking Personal Data. The Wellcome Trust Limited. London.

Westin, A. F. 1967. *Privacy and freedom,* New York, Atheneum.

World Economic Forum 2013. Unlocking the Value of Personal Data: From Collection to Usage. *Retrieved April 1, 2014, from World Economic Forum Website.* WEF.

Zhu, F., Carpenter, S., Kulkarni, A., Chidambaram, C. & Pathak, S. 2009. Understanding and minimizing identity exposure in ubiquitous computing environments. *In Mobile and Ubiquitous Systems: Networking & Services, MobiQuitous, 2009. MobiQuitous' 09. 6th Annual International. IEEE***,** 1-10.