The Effects of Different Personal Data Categories on Information Privacy Concern and Disclosure

Keywords
Personal Data Categorization; Information Disclosure; Privacy Concern; Information Privacy; Privacy by Design

Abstract
The potential threats of exposing personal data associated with online services have been a reason for concern, and individuals as customers may decline to disclose their data due to trust issues. Literatures have shown evidence that greater transparency in the types and purposes of data requested encourages individuals to disclose personal data. This evidence indicates a need to examine the characteristics of personal information practices. Furthermore, current legislations recognize the presence of different data characteristics such as location-specific, health-specific and financial-specific. Yet, current legislations are formed to identify personal data as a singular category regardless of the requirements, including the specification of processed personal data to be relevant and limited to what is necessary for the purposes of enabling functional services. Without categorization, measuring "relevant" and "necessary" can be ambiguous. Several researches have explored the impact of personal information type and sensitivity level on privacy concern and disclosure; however, most of them lacked an in-depth examination of data categorization with systematic validation. Our study aims to fill this gap, and additionally further look into how contextual demographic factors influence the perception on information privacy concern and disclosure of different personal data categories from a Malaysian perspective. Our study provides new evidence of validated personal data categories and their significant differences in perceived information privacy concern and disclosure intention. Our research finding also discovers that Age, Gender and Working Industry, as demographic factors, have significant effects on disclosure intention associated with Tracking, Finance, Authenticating and Medical-health information.

1. Introduction

Today's internet network capability in providing bigger bandwidth and faster data transfer speed has facilitated a conducive environment for individuals to use online services as well as store information in the cloud. Privacy threats associated with online application services have long been a reason for concern and individuals as service users or customers may even decline to disclose their personal data due to privacy trust issues (Wang & Peng, 2013). Organizations capitalize on customer data in order to gain competitive advantage against others (Janssen & van den Hoven, 2015; TRUSTe, 2011). By demanding irrelevant and loosely defined permissions of users to disclose personal data in exchange for service provisioning, plus even with highly personalized data aggregation, application service providers could be in the position to provide third parties with sensitive data (Enck et al., 2014). Despite this possibility, most individuals as service users are unable to understand the technical mechanisms of how their personal data is being processed in cases of data leakage (Acquisti et al., 2016).

Regarding information-privacy related behaviour, various studies have shown the existence of users' conflicting privacy-paradox attitude towards their privacy concerns and actual behaviour. Individuals as application service users who demonstrate concern about their information privacy however perform little action in protecting their personal data (Norberg, 2007; Barth & De Jong, 2017). Despite the privacy-paradox attitude, the problem of privacy trust is rising rapidly due to unauthorized sharing of personal data and increasing cases of data leakage (Cradock et al., 2017). The presence of transparency in how personal data will be processed and used could build an individual's confidence and trust towards an organization. As a prior study has shown, in order to improve transparency, organizations need to inform users on what personal data is being collected and how it is being used (Cradock et al., 2017). Literatures also show that greater transparency in terms of the types of data requested (Phelps et al., 2000; Park et al., 2018) and the purposes for their use (Anderson & Agarwal, 2011) positively impacts an individual's beliefs of service providers' practices that could influence the former's concern and disclosure attitude. Existing research (Malhotra et al., 2004; Bansal & Gefen, 2010; Milne et al., 2017) also found that generally requests for more sensitive information decrease trust and disclosure intention, which indicates a necessity to design studies to examine and differentiate characteristics of personal information.

While there is no fixed definition of transparency, it carries the responsibility of organizations as data controllers to notify users on how individual personal data is being used or processed as required by most data protection regulations (GDPR, 2018; PDPA, 2013). Despite the importance and the requirements for transparency in data collection as indicated by prior studies and regulations, the current legislations are built to recognize personal data as a singular category regardless of the presence of different characteristics such as location-specific, health-specific, and financial-specific. Standard for measurement can be vague and subjective in realising transparency without data categorization. For example, the European General Data Protection Regulation (GDPR, 2018) requires that "the processed personal data must be adequate, relevant and limited to what is necessary for the purposes for which it is processed". Measurement of "adequate", "relevant" and "necessary" can be ambiguous without the understanding of personal data characteristics, and the magnitude of different characteristics' impact on individuals' concern over the potential threat of disclosing the data. Several prior studies (Rumbold & Pierciokne k, 2018; Robinson, 2016; Anderson & Agarwal, 2011; Bansal &

Gefen, 2010; Malhotra et al., 2004) have investigated the effect of information type and sensitivity level on privacy concern and disclosure; however, these studies did not focus on validating the differences between personal data categories. A study conducted by Phelps et al. (2000) investigated privacy concern and willingness to disclose personal information, including the examination of types of information (i.e. personal finances, media habits, lifestyle and demographics) and how they affect consumers' concern and likelihood of purchase. As Phelps et al.'s work was carried out almost twenty years ago, more and new varieties of information types have emerged due to the evolution of internet technology, for instance location-related tracking data and social media posts that lead to the exposure of behavioural information. Moreover, a thorough study on personal data categorization with validity assessment was not included in Phelps et al.'s research and other prior studies.

While an extensive study by Milne et al. (2017) and Park et al. (2018) presented types of information associated with risk and data value respectively, we consider a further study necessary to understand how different personal data categories are perceived in relation to privacy concern along with disclosure intention. In addition, a prior study by Chua et al. (2018) showed evidence that demographic factors bring an impact to employees' information security awareness and compliance behaviour. This evidence motivates us to extend the understanding by investigating how demographic factors influence individuals' perceptions on personal data categories through our study in the Malaysian context.

With this motivation, our research study seeks to answer the following questions: (i) What are the valid personal data categories based on the nature of their characteristics and the findings of prior studies? Consecutively, (ii) Are these different data categories perceived with the same level of importance in relation to information privacy concern and the disclosure intention? (iii) In comparison, how differently are information privacy concern and disclosure intention perceived? (iv) How do the demographic factors influence perceived information privacy concern and disclosure intention for the different data categories? To answer the first research question, we conducted a qualitative examination to identify personal data categories with a validity test. After testing the validity of personal data categories, we answered the second and third questions through the following statistical hypotheses:

Hypothesis 1 (H1). Different personal data categories have different perceived importance levels of information privacy concern.

Hypothesis 2 (H2). Different personal data categories have different perceived importance levels of disclosure intention.

We subsequently performed statistical tests to derive the answer for research question four.

2. Literature Review

2.1 Personal Data and Regulatory Protection

Personal data refers to any information that may relate to an individual and may be used as an identification, particularly by reference to an identifier such as a name, place data, online identifier and identity number, or to one or more variables specific to that individual (GDPR, 2018; Milne et al., 2017). In other words, personal data consists of any information about living persons that may be identified by the data or from combinations of data and other information possessed or likely to be possessed by the person in control of the data.

Due to the possibility of personal data violations and misuse, individuals see personal data security as their major concern when performing online activities (Tsai et al., 2011). To address the concerns of personal data violation and to balance the interests of individuals and organizations, various international guidelines exist, while country-specific regulations are enforced to govern appropriate data collection and use. Examples of commonly mentioned laws include the European Union's General Data Protection Regulation (GDPR), the California Consumer Protection Act (CCPA), the US Federal Trade Commission (FTC)'s Fair Information Practices Principles (FIPPs), the Freedom of Information Act 2000 (FIA), and the Organization for Economic Cooperation and Development (OECD) guidelines.

The General Data Protection Regulation (GDPR) enforced by the European Union (EU) is considered one of the tightest data protection laws to-date from the aspects of worldwide data coverage and penalty. The GDPR's primary aim is to significantly enhance individual data protection rights, ensure free flow of personal data on the digital market, boost transparency, and decrease administrative burden (GDPR, 2018). The EU has introduced new requirements on what organizations as data controllers may need to be transparent in with regards to the categories (i.e. types) of personal data they process. However, it remains uncertain what kind of personal data might fall into the personal data category. Hence, consideration for a new approach towards the protection of each personal data category is needed in order to enhance transparency, allowing different levels of protection to be imposed on different types of personal data categories.

2.2 The Value of Personal Data

The privacy of personal data is generally a state of restricted access to the personal information of a person. Personal data should be secured because it carries financial merit in this data-driven economy, as it can be disclosed by individuals in exchange for incentives in the form of free digital facilities or for product or service discounts (Sidgman & Crompton, 2016). In other words, personal data can be used in the digital economy in return for digital content instead of cash

(Malgieri & Custers, 2017). However, most people are not conscious of their personal data's financial value. If people were shown the financial value of their personal data, they may gain a greater level of awareness of their power of control over their personal data and make the correct choices before disclosing their personal data (Malgieri & Custers, 2017).

The value of data often lies in relation with other data, generating new information. Data collected, aggregated and processed appropriately can help organizations to better comprehend customers' behaviours and preferences (Chang et al., 2018). When used correctly, these data are valuable in conferring businesses the competitive advantage in providing product/service customization and personalization (Erevelles et al., 2016).

2.3 Importance of Personal Data Categorization

Cradock et al. (2017) demonstrated how varied the sources from scholars to laws and privacy professionals have categorized and grouped personal data. There are multiple parties concerned, which often differ in terms of their granularities, who distinguish between personal data by defining "kinds", "categories" and "objects". Therefore, the requirement for service providers, as data controllers, to inform individuals, as data subjects, of the categories of personal data that is being processed could be interpreted differently in practice in the absence of scientific study to provide further guidance.

A group of individuals or entities with common features can be defined as a "category" (Soanes, 2011). Categorizing personal data enables an individual to find out what "things" are, simply because he/she knows which category they belong to (Hunn, 1979). Categorising personal data could also increase transparency in the processing of personal data, by understanding which category of personal data is being handled, and using the category as a connecting anchor for additional information. Moreover, knowing the differences between categories of personal data allows businesses or data controllers to evaluate potential threats in their data processing (Milne et al., 2017; Cradock et al., 2017). This is critical for data controllers when operating a data protection impact evaluation (Vollmer, 2018). Consequently, the evaluation can be used as a guideline on what organizational and technical measures are needed to enhance the security of personal data. Setting up different levels of security based on categories of personal data will be very costly for organizations. Thus, organizations need to have a profound understanding of the different categories of personal data that they process in order to understand and identify the appropriate technical and organisational measures necessary.

Data monetization is gradually becoming an issue of concern in European legislation and therefore it is necessary to ensure a future-proof protection of consumer data. Regulations imposed on digital content provided in exchange of personal data can be a factor to raise consumer awareness of the financial importance of their personal data, thereby leading to better protection of their personal data (Malgieri & Custers, 2017). From a consumer perspective, transparency in today's big data generation is crucial in gaining their trust (Cradock et al., 2017). Categorization of personal data allows the reduction of the amount of data needed to be provided in order to improve transparency (Wang & Peng, 2013).

2.4 Information Privacy Concern

Information privacy can be defined as the ability of an individual to control his/her personal information whereas information privacy concern is denoted as an individual's concern about organizational practice related to the collection and use of personal information (Smith et al., 1996). Though personal data can be used by businesses to personalize product/service provision and by individuals to exchange for incentives/services/products, data handling remains a concern for individuals and this is further exacerbated by the rise of data leaks. Given that each piece of data leaves behind electronic trails of customer activities, individuals are concerned about how companies collect and use their private information (Janssen & Kuk, 2016; Morey et al., 2015).

Individuals with information privacy concern protect their privacy by reacting negatively to organizational information practices when they perceive their privacy rights being threatened (Smith et al., 1996). For organizations that operate their business in an online environment, information privacy is a critical ethical issue since organizations reply on their capability to collect huge amounts of personal information (Son & Kim, 2008) as customer data is an asset when organizations utilize them strategically. Therefore, securing customer data to address customer concern should be an organization's priority in order to establish customer trust; this should come prior to the organization's plan to leverage on customer data (Chua et al., 2018) in order to prevent cases of data breaches that could eventually tarnish an organization's image. Together with the growing amount of internet data leaks (Wang & Peng, 2013), these incidents increase the concerns about customer privacy towards data danger (Drinkwater, 2016).

Consequently, the development of personal data protection policies governing the management and security of personal data is essential for balancing customers' privacy concerns and the organizations' obligation to strategize client data for their company benefit. Considering the potential risks and losses, governments are enforcing regulations and policies (such as GDPR, FIPPs, FTC and CCPA) on privacy to safeguard individuals from potential detrimental acts.

2.5 Personal Information Disclosure Behaviour

Expectancy theory suggests that an individual considers the overall possible consequences and seeks to minimize negative consequences and maximize positive consequences in his/her motivation to act or not (Vroom, 1964). In the context of information privacy research, individuals weigh costs and benefits when determining if they were going to disclose personal information (Culnan & Armstrong, 1999). Dinev and Hart (2006) discovered that a higher level of perceived information privacy concerns yields to a lower level of willingness to disclose personal information. Research suggests that individuals can be motivated to disclose their personal information when higher levels of trust in an organization exist (McKnight et al. 2002) as well as when they are aware of how an organization uses and manages their personal information (Culnan & Armstrong, 1999).

Communication privacy management (CPM) theory generally supports that individuals make decisions of personal information disclosure based on the contextual criteria they perceive at the time the decision must be made (Petronio, 2002). In this context, there are risks associated with personal information disclosure (e.g. personal information misused or transferred by organizations to third parties) but also potential benefits (e.g. service provisioning in relationships between consumers and organizations).

Increasing use of technology, especially on the internet, has fuelled the requirement of users disclosing their personal data online in exchange for application services. Disclosure of personal data is a prerequisite for customers accessing services or making online purchases, or when organizations provide certain customized services to meet the needs of customers. The increasingly social nature of many web-based social network sites also places a price of privacy on users due to an increased necessity to disclose personal data as part of system functionalities (Joinson, 2008; Ahern et al., 2007).

Besides exchanging personal data for access to services, customers could also be provided a discount on the total service cost to encourage them to reveal their personal data. Such monetary benefit could, for example, be in the form of a digital wallet provided by the organization to encourage customers to reveal their personal data (Malgieri & Custers, 2017). Providing personal data for personalization services can be another reason of disclosure. Customers are urged to reveal their personal data in order to gain a more customized service, such as a personalized search engine or a customized social network platform. In some cases, the online services offered may lose certain functionalities when they could not be personalized (Malgieri & Custers, 2017).

Initial feelings created from a general impression of a website before exchanging data may differ from those experienced at a later point when internet customers evaluate the exchange of data based on the price, benefit and perceived fairness of a social agreement (Li et al., 2011). In regards to social network, Dwyer et al. (2007) found that the more users trust a website, the more willing they are to disclose information and develop contacts on these social network sites (Dwyer et al., 2007; Wang & Peng, 2013).

An individual behavioural intention to disclose personal data can also be affected by his or her belief (Hausenblas et al., 1997) – for example, an individual's belief that using a location-tracking application could give rise to both positive and negative effects, with the positive being the pinpointing of the needed location, and the negative being the service provider's knowledge of where he or she frequents, which might be dangerous.

2.6 Related Works on Personal Data Categorization

Categorizing personal data enables greater transparency by allowing individuals to gain more information about the category of personal data being processed, ultimately building customers' confidence in disclosing their personal data. This rationale is supported by prior studies showing that greater transparency, in terms of the types of data requested (Phelps et al., 2000) and the purposes for their use (Anderson & Agarwal, 2011), positively impacts an individual's beliefs in service providers' practices that could eventually influence the former's concern and disclosure attitude. Studies (Bansal & Gefen, 2010; Malhotra et al., 2004; Milne et al., 2017) also found that the more sensitive a piece of information, the lower the disclosure intention, which indicates a necessity to design studies to examine and differentiate characteristics of personal information.

Prior studies by Rumbald and Pierscionek (2018), Anderson and Agarwal (2011), Bansal and Gefen (2010), and Malhotra et al. (2004) investigated the effect of information type and sensitivity level on privacy concern and disclosure; however, these studies did not demonstrate the validity of the differences of data categories. On the other hand, Phelps et al. (2000) presented a study on types of information including personal finances, media habits, lifestyle and demographics and how they affect consumers' concern and likelihood of purchase. Yet, we argue that Phelps et al.'s (2000) list of information categories can be further extended due to the evolution of internet technology in the past two decades, leading to more and new varieties of data categories, for instance location-related tracking data and social media posts that lead to the exposure of behavioural information. Further, a more thorough study on personal data categorization with validation analyses were either not the focus or excluded in Phelps et al.'s and prior studies.

Several technology patents deploy the method of using different personal information categories for processing data. Degele et al. (2017) proposed a data model and application architecture for a digitized health insurance, using a predefined personal information categorization of fitness profiles (such as pulse rate, heart rate, distance covered, and number of steps); contact information (address, email and telephone); identity information (name and birthdate); and device used tracking. Brannon et al. (2020) presented an automated system to score the sensitivity level of text-based documents by way of breaking up pieces of information and assessing their sensitivity score based on the personal data classification predefined in the system. The automated system pre-classifies personal information into Personal Identifiable Information (PII) (such as contact details, addresses, job related information, full name, birthdate, marital status, employment status, employee information such as tax identification, social security and user account numbers); Partial PII (first name or last name, gender, zip code or street or state or city or country, marital status, employer name); and Sensitive PII (employment status, marital status, user account number, social security number, tax identification number, health insurance details, health plan account number, employer identification number). Muffat and Kodliuk (2020) proposed a system to extract information entities from text and predict the likelihood of those entities as PII based on the system's predefined classification surrounding the business customer context such as first name, last name, salutation, client business relationship, cash account numbers, custody account numbers, portfolio ID, contract number for e-banking, phone number, address, credit card number, company name, passport number. Systems presented through these technology patents (Degelete et al., 2017; Brannon et al., 2020; Muffat & Kodliuk, 2020) did not report any significant tests to validate the differences between personal data categories.

Park et al. (2018) examined the perceived value of personal information types based on responses from 44 Korean female participants. The personal information categories identified in this study were health information, social information, financial information, online information and demographic information. The Analytic Hierarchy Process (AHP) was applied to validate the results. Milne et al. (2017) studied and ranked 52 information types along with four perceived risk categories (physical, psychological, monetary and social), information sensitivity and willingness to provide. Personal information categories by customer segments identified in this study were basic demographics, personal preferences, contact, community interaction, financial information and secure identifier.

3. Research Methodology

3.1 Research Design

In order to gain insight into how different personal data categories vary in terms of their perceived information privacy concern and disclosure intention, we conducted a survey in Malaysia for our experimental study. All personal identifiable information was not collected, and responses remained anonymous.

3.2 Survey

A structured questionnaire was designed, and five-point (1-5) Likert-type questions were used. There were three sections in this survey. The first section contained questions related to demographic information. The other two sections comprised questions that required the rating of perceived information privacy concern and disclosure intention based on different personal data categories respectively. As Malaysia is a multi-racial country, the questionnaires were prepared and made available in three languages, namely, English, Malay and Chinese. The appendix at the end of this paper presents the questionnaire in English.

3.3 Data Collection and Sample Size

The self-administered questionnaire was created through an online data collection service provider, namely, the SurveyMonkey platform. A total of 465 valid responses were collected within the span of two months, from the beginning of January till end of February 2020. The selection criteria of this survey determined that only Malaysians aged 18 or above were qualified to respond to the questionnaire.

3.4 Classification of Personal Data Categories

As part of the items asked in the questionnaire, we first examined categories of personal data based on prior studies, then aggregated and classified them into six categories. The classification of the categories was adapted from the literatures of Phelps et al. (2000), Milne et al. (2017), Park et al. (2018) and Rumbold and Pierscionek (2018). The categories from these literatures were further aggregated based on the nature of their characteristics as the criteria for consideration. Table 1 presents how the six data categories were derived based on the aggregation.

Table 1: Categories and Characteristics of Personal Data

Present Research	**Phelps et al. (2000)**	**Robinson (2016)**	**Milne et al. (2017)**	**Park et al. (2018)**	**Rumbold & Piercioknek (2018)**
Social-economic	Demographic data	Demographic data Work-related information	Basic demographics	Demographics	Socio-economic (Human demographics) Readily apparent human characteristics (protected and unprotected)

Lifestyle-behavior	Lifestyle interaction Media habits	Life history information	Personal preferences Community interaction	Social information	Human-machine interactions (browsing history/logs) Socio-economic (Human behaviour, thoughts and opinions)
Tracking	*(Not mentioned)*	Contact information	Contact information	Online information	Human-machine interactions (device tracking)
Financial	Financial data	*Payment information*	Finance information	Finance information	*(Not mentioned)*
Authenticating	Personal identification data	Online account information	Secure identifier	*(Not mentioned)*	*(Not mentioned)*
Medical-health	*(Not mentioned)*	*Medical history*	*(Not mentioned)*	Health information	Medical or healthcare data
(Not applicable)	*(Not applicable)*	*Not applicable)*	*(Not applicable)*	*(Not applicable)*	Non-personal data *(NOTE: not applicable - we consider this category beyond the context of our study)*

Cohen's κ test (McHugh, 2012) was run to determine if there was an agreement between the three researchers' judgment, that is, whether the list of categories aggregated in Table 1 and the items (i.e. characteristics) associated with them as presented in Table 2 are valid according to the nature of the data characteristics.

Table 2: Categories and Characteristics of Personal Data

Category	**Characteristics**
Lifestyle-behaviour (LB)	Information about an individual's lifestyle and characteristics that influence his/her relationship or community connection, preferences, habits, beliefs or opinion. Examples: LB1. Belief (e.g. religious beliefs, philosophical beliefs, thoughts, etc.) LB2. Preferences or interests (e.g. opinions, intentions, interests, favorite foods, colors, likes/dislikes, etc.) LB3. Behavior (e.g. browsing habit, call patterns, links clicked, demeanor, attitude, etc.) LB4. Family/relationship (e.g. family structure, siblings, offspring, marriages, divorces, relationships, etc.)
Social-economic (SE)	Information that describes an individual's social demographics or status or information that reflects those characteristics. Examples: SE1. Ethnicity (e.g. race, national/ethnic origin, languages spoken, dialects, accents, etc.) SE2. Physical characteristics (e.g. picture, video, etc.) SE3. Demographics (e.g. age, gender, etc.) SE4. Professional career (e.g. job titles, salary, work history, schools attended, employment history, etc.)
Tracking (T)	Information that provides a mechanism for locating and contacting an individual. Examples: T1. Contact information (e.g. email address, physical address, telephone number, etc.) T2. Communication (e.g. telephone recordings, voice mail, text messages, etc.) T3. Location (e.g. country, GPS coordinates, room number, etc.) T4. Computer device details (e.g. IP address, Mac address, browser information, digital fingerprints, etc.)
Financial (F)	Information that identifies an individual's income, financial account, credit, purchasing/spending capacity, and assets owned/rented/borrowed/possessed. Examples: F1. Credit history (e.g. credit records, credit worthiness, credit standing, credit capacity, etc.) F2. Assets (e.g. property, personal belongings, etc.) F3. Financial account (e.g. credit card number, bank account, etc.) F4. Transactions (e.g. purchases, sales, credit, income, loan records, transactions, taxes, purchases and spending capacity, etc.)
Authenticating (A)	Information used to authenticate an individual. Examples: A1. Passwords or pin (e.g. bank account password or pin, email address password, etc.) A2. Identity code (e.g. government issued identification, etc.) A3. Username (e.g. social media username, online banking username, etc.)
Medical-health (MH)	Medical conditions or health-related information of an individual. Examples: MH1. Diagnoses (e.g. test results, health records, prescriptions, physical and mental health, disabilities, etc.) MH2. Genetic data (e.g. genetic information, blood type, etc.) MH3. Personal health history and medication experiences

3.5 Pilot Testing

Before the distribution of the finalized questionnaire to the respondents, a pilot study was carried out using a small sample of five to evaluate the clarity of the questions. The five participants involved in the pilot test were made up of three males and two females with a combination of occupations that included employment in the private sector and the health sector as well as student, and age ranging from 22 to 58 with a mean/median of 37.2/26. The feedback from the pilot test was overall satisfactory in terms of understanding of the questionnaire requirements and content. The only revision made based on the feedback was to condense some relevant items of the questionnaire to address the comment stating that the questionnaire was too long.

3.6 Confirmatory Factor Analyses

The survey questions based on the characteristics proposed in Table 2 were composed of different dimensions. Each item of the construct was intended to address one of the six dimensions of the 'Personal Data Categories': Lifestyle-behaviour, Social-economic, Tracking, Financial and Authenticating. To test the validity and reliability of the questionnaire and their fit in the respective data category, a component factor analysis method was performed.

The factor loading represented the relative perceived importance of each *item* (i.e. each question of our questionnaire such as A1, A2 and A3) related to each *factor* (i.e. data category such as "Authenticating").

Cronbach alpha (α) was used to assess the average measure of internal consistency and item reliability, whereas Composite reliability (CR or sometimes called construct reliability) was used to measure the scale reliability in overall for a factor with minimum threshold of 0.7 for both α and CR (Brunner & Süβ, 2005; Hair et al., 2009). To assess the internal reliability, the Cronbach's coefficient α is calculated (Cronbach, 1951). With a set of *i* items $\lambda_1, \lambda_2, \ldots, \lambda_i (i \geq 2)$ composing the composite $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_i$, we have α defined as:

$$\alpha = \frac{i}{i-1}\left[\frac{\sum_{1 \leq a \neq b \leq k} \mathrm{Cov}(\lambda_a, \lambda_b)}{\mathrm{Var}(\lambda)}\right]$$

The variables Cov and Var denote covariance and variance, respectively, and $1 \leq a \neq b \leq k$ stands for all possible inter-item covariances.

An exploratory factor analysis was performed to evaluate the validity of the construct that measures the individual 'Personal Data Categories' dimensions (Swisher et al., 2004). To establish discriminant validity, an average variance extracted (AVE) analysis was performed (Bertea & Zait, 2011). The formula to calculate the value of Construct Reliability (CR) and Average Variance Extracted (AVE) are shown below:

$$CR = \frac{\sum_{i=1}^{k} \lambda_i^2}{\sum_{i=1}^{k} \lambda_i^2 + \sum_{i=1}^{k} Var(e_i)}$$

$$AVE = \frac{1}{k}\sum_{i=1}^{k} \lambda_i^2$$

The variable k represents the number of items in λ_i the factor loading of item i and $Var(e_i)$ denotes the variance of the error of item i.

Average Variance Extracted (AVE) was used as a measure of the amount of variance that is captured by a factor in relation to the amount of variance due to measurement error indicating how well the items in a factor can correlate with one another. The value of AVE for a factor should meet a suggested critical value of 0.50 or above (Fornell & Larcker, 1981).

3.7 Data Analysis Methods

There are several approaches for this methodology. Firstly, the Shapiro-Wilk test (Shapiro & Wilk, 1965) was used to the normality of data to determine whether our sample data had been drawn from a normally distributed population (Yap & Sim, 2011). The data were found not normally distributed, so the non-parametric Friedman test (Conover, 1998) was used to compare the privacy concern and disclosure intention based on the ratings between multiple categories of personal data. For descriptive data analysis, summary statistics were generated in order to obtain the median, interquartile range in understanding the age of the respondents. Whereas for categorical data (including nominal and ordinal data), percentages and frequencies were presented for descriptive analysis.

For hypothesis testing, the Friedman test was carried out to assess if there were statistically significant differences in levels of perceived importance among different personal data categories in terms of information privacy concern and disclosure intention.

Mean rank was used to compare the differences in the scores of the data categories. Mean rank was used because the distributions for each category were different. The mean rank value of each category provides an understanding of how much a given category tends to have high values. In other words, if the mean rank for a category is smaller than that of the other, this indicates that the median for the category is most likely smaller than the other. To compare if two data categories were statistically significantly different, post-hoc pairwise comparisons using Wilcoxon test (Derrick and White, 2017) was conducted. Because post-hoc tests are used to confirm the differences occurring between personal data categories, they were only run when we observed an overall statistically significant difference in group means using the adjusted Bonferroni *p*-value.

For group comparisons of the demographic characteristics on perceived privacy concern and disclosure intention associated with different personal data categories, our selection of analysis methods was based on the following rationale:

- The data collected from the respondents for the importance of the different data categories were 5-point Likert scale data, thus non-parametric statistical tests were used.
- The independent variable Gender consisted of two groups (Male or Female), hence the Mann-Whitney U test (Mcknight & Najab, 2010a) was used.
- All the other independent variables (Age, Race, Working Industry) consisted of more than two groups, therefore the Kruskal-Wallis H test (Mcknight & Najab, 2010) was used.
- For the Mann-Whitney U and Kruskal-Wallis H tests, we used the median (Zhang & Zhang, 2009) of demographic groups to compare the respondents' perceived level of privacy concern and disclosure intention.

3.8 Data Preparation

There was no missing data found and no removal of incomplete data from the data collected. To obtain the level of privacy concern and disclosure intention of each category, data transformation using the method of deriving the mean values was carried out to calculate the average level of disclosure intention and privacy concern.

4.0 Results

4.1 Demographic Analysis

Table 3 below shows the descriptive analysis of the respondents who were involved in the research survey.

Table 3: Demographics of the Sample (N = 465)

Age – Median (IQR)		36 (18 – 60)
Age group – N (%)	Below 25	85 (18%)
	25-29	62 (13%)
	30-34	63 (14%)
	35-39	59 (13%)
	40-44	76 (16%)
	45-49	26 (6%)
	50-54	46 (10%)
	55 and above	48 (10%)
Occupation (industry) – N (%)	Others	82 (17.2%)
	Architecture/Engineering/Real estate/Transportation/Utilities/Wholesale	75 (16.1%)
	Private employment	66 (14.2%)
	Direct selling/retailer	39 (8.4%)
	Student	38 (8.2%)
	Audit/Accountancy/Legal	36 (7.7%)
	Education	32 (6.9%)
	Banking/finance	26 (5.6%)
	Health/Insurance	25 (5.4%)
	Telecommunication	23 (4.9%)
	Government agencies	13 (2.8%)
	Tourism/Hospitality	10 (2.2%)
Gender – N (%)	Male	255 (54.8%)
	Female	210 (45.2%)
Race – N (%)	Malaysian Chinese	217 (46.7%)
	Malaysian Malay	189 (40.6%)
	Malaysian Indian	59 (12.7%)

A total of 465 eligible responses were collected within the span of two months. The mean and median age of the respondents were 37.2 and 36 respectively, showing a considerably balanced distribution whereby there was no age group extremely dominating the sample. Similarly, with the occupation factor, we observed no dominance among the industries. Respondent genders were almost equally distributed, with female 54.8% and male 45.2%. We observed that most of our respondents were Malaysian Malay and Malaysian Chinese, which totalled 406 respondents (87.3%), followed by Malaysian Indian. As there was no data available showing specifically the population distribution between the ages 18 and 60 from the Malaysian Statistics Department, it was unfeasible to derive and confirm the statistical significance of the Race balance in ratio of Malaysian population.

4.2 Validation of Personal Data Categories and Their Associated Characteristics

For ensuring the validity of the personal data categorization process aggregated in Table 1, Cohen's κ test was performed to determine if there was an agreement between the three researchers' judgment on whether the list of categories and the items (i.e. characteristics) associated with them as presented in Table 2 are valid according the data characteristics. There was perfect agreement between the three researchers' judgements, $\kappa = 1.000$ (95% CI, .300 to .886), $p < .0005$. The Cohen Kappa coefficient (κ) represents a statistical measure of inter-rater reliability that is used to determine the agreement between three researchers, which κ value < 0 indicates no agreement, 0–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1 as almost perfect agreement.

4.3 Results of the Component Factor Analysis

Factor loadings, Cronbach's Alpha, Average Variance Extracted, and Composite Reliability were determined to assess the reliability and validity of the personal data categories. Table 4 shows the factor analysis and reliability test for this study.

Table 4: Factor Analysis and Reliability Test

	Disclosure Intention				**Privacy Concern**			
Personal Data Category	**Factor Loading**	α	**AVE**	**CR**	**Factor Loading**	α	**AVE**	**CR**
Lifestyle-behaviour		0.71	0.54	0.82		0.71	0.54	0.82
LB1	0.71				0.78			
LB2	0.76				0.76			
LB3	0.70				0.73			
LB4	0.75				0.66			
Social-economic		0.76	0.60	0.84		0.76	0.58	0.85
SE1	0.80				0.79			
SE2	0.70				0.73			
SE3	0.81				0.81			
SE4	0.72				0.73			
Tracking		0.79	0.61	0.86		0.79	0.61	0.86
T1	0.76				0.805			
T2	0.71				0.729			
T3	0.85				0.762			
T4	0.81				0.834			
Financial		0.89	0.75	0.92		0.86	0.71	0.91
F1	0.86				0.84			
F2	0.85				0.83			
F3	0.89				0.84			
F4	0.86				0.86			
Authenticating		0.80	0.71	0.88		0.78	0.70	0.87
A1	0.87				0.88			
A2	0.80				0.82			
A3	0.85				0.81			
Medical-health		0.81	0.74	0.89		0.82	0.75	0.90
MH1	0.92				0.95			
MH2	0.71				0.68			
MH3	0.93				0.95			

The findings of the CFA confirm that most of the factor loadings were above 0.7, meeting the minimum acceptance threshold of 0.7, with the exception of LB4 (0.66) and MH2 (0.68) having factor loading value slightly less than 0.7 for the "Privacy concern".

Our test results showed the α and CR of all data categories as being above the threshold with values ≥0.7 and ≥0.8 respectively, indicating all the items in their respective data category as consistent and reliable.

The AVE results we obtained showed that all data categories exceeded 0.50, implying that all the questionnaire items in their respective data category correlated well with one another.

4.4 Hypothesis Test

Hypothesis 1 (H1). This hypothesis is supported

The Friedman test was carried out and results showed that there were statistically significant differences in levels of perceived importance between different personal data categories in terms of privacy concern ($\chi^2(5) = 480.3$, $p < 0.001$). The mean rank for each personal data category is shown in Table 5.

Table 5: Mean Rank for Privacy Concern by Category of Personal Data (N=465)

Category	Mean Rank
Authenticating	4.45
Finance	4.30
Tracking	3.81
Medical-health	3.12
Lifestyle-behaviour	2.66
Social-economic	2.65

Based on the category mean ranks presented in Table 5, it can be implied that the respondents had the highest privacy concern with regards to the Authenticating category of their personal data, followed by Finance; both showed no statistically significant difference in their mean ranks based on the Pairwise comparison result as presented in Table 6. The lowest privacy concern for the respondents were the Lifestyle-behavior and Social-economic categories. Both Lifestyle-behavior and Social-economic categories also posed no statistically significant difference in their mean ranks. Medical-health information scored a nearly neutral concern level.

The post-hoc pairwise comparisons using Wilcoxon test in Table 6 demonstrated that there was no significant difference in the comparisons between the “Lifestyle behavior – Social-economic” categories and between the “Finance – Authenticating” categories. In the context of this study, the Z score shows how far away two data categories are from the mean relatively. The Z score is positive if the value lies above the mean, and negative if it lies below the mean.

Table 6: Pairwise Comparisons for the ‘Concern’ Factors

Comparisons	Z score	Adjusted *p*-value
Lifestyle-behaviour – Social-economic	0.105	1.000
Lifestyle-behaviour – Tracking	-9.33	<0.001*
Lifestyle-behaviour – Finance	-13.32	<0.001*
Lifestyle-behaviour – Authenticating	-14.59	<0.001*
Lifestyle-behaviour – Medical-health	-3.72	0.003*
Social-economic – Tracking	-9.44	<0.001*
Social-economic – Finance	-13.43	<0.001*
Social-economic – Authenticating	-14.70	<0.001*
Social-economic – Medical-health	-3.82	0.002*
Tracking – Finance	-4.00	0.001*
Tracking – Authenticating	-5.26	<0.001*
Medical-health – Tracking	5.62	<0.001*
Finance – Authenticating	-1.27	1.000
Medical-health – Finance	9.60	<0.001*
Medical-health – Authenticating	10.88	<0.001*

* Mean rank comparison is significantly different

Hypothesis 2 (H2). This hypothesis is supported.

The Friedman test result showed that there was a statistically significant difference in level of perceived importance among the different personal data categories in terms of disclosure intention ($\chi^2(5) = 559.6,\ p < 0.001$). The mean ranks for each personal data category are shown in Table 7.

Based on the mean rank and the significant differences proven in the pairwise comparisons results presented in Table 7 and Table 8 respectively, it can be inferred that the Financial category of personal data was the least likely personal category to be disclosed by the respondents, followed by Authenticating, Tracking, Medical-health, Lifestyle-behavior and Social-economic.

Table 7: Mean rank for disclosure intention by category of personal data (N=465)

Category	**Mean Rank**
Social-economic	4.55
Lifestyle-behaviour	4.37
Medical-health	3.67
Tracking	3.37
Authenticating	2.58
Finance	2.45

Table 8 presents the post-hoc pairwise comparisons using the Wilcoxon test for the ‘disclosure’ factors, and confirms that all pairwise comparisons between categories showed significant mean rank differences, except for the “Lifestyle-behavior – Social-economic”, “Tracking – Medical-health”, and “Finance – Authenticating” pairs.

Table 8: Pairwise Comparisons for the ‘Disclosure’ Factors

Comparisons	**Z score**	**Adjusted *p*-value**
Lifestyle-behaviour – Social-economic	-1.42	1.000
Tracking – Lifestyle-behaviour	8.15	<0.001*
Finance – Lifestyle-behaviour	15.64	<0.001*
Authenticating – Lifestyle-behaviour	14.61	<0.001*
Medical-health – Lifestyle-behaviour	5.71	<0.001*
Tracking – Social-economic	9.57	<0.001*
Finance – Social-economic	17.06	<0.001*
Authenticating – Social-economic	16.03	<0.001*
Medical-health – Social-economic	7.13	<0.001*
Finance – Tracking	7.49	<0.001*
Authenticating – Tracking	6.46	<0.001*
Tracking – Medical-health	-2.44	0.223
Finance – Authenticating	-1.03	1.000
Medical-health – Finance	-9.93	<0.001*
Authenticating – Medical-health	-8.90	<0.001*

* Mean rank comparison is significantly different

4.5 Demographics Analysis Associated with Different Personal Data Categories

Table 9 and 10 show the group comparison results of demographic characteristics on disclosure intention and perceived privacy concern associated with different personal data categories respectively.

Table 9. Demographics Differences in Disclosure Intention Associated with Personal Data Categories

		Lifestyle-behaviour	**Social-economic**	**Tracking**	**Finance**	**Authenticating**	**Medical-health**
	N	**Test/p-value Median**	**Test/p-value Median**	**Test/p-value Median**	**Test/p-value Median**	**Test/p-value Median**	**Test/p-value Median**
Age		$\chi^2(7)$ = 5.045 p = 0.654	$\chi^2(7)$ = 3.673 p = 0.817	$\chi^2(7)$ = 17.561 p = 0.014*	$\chi^2(7)$ = 22.102 p = 0.002*	$\chi^2(7)$ = 21.020 p = 0.004*	$\chi^2(7)$ = 25.927 p = 0.001*
Below 25	85	3.000	3.250	2.000	1.250	1.667	2.333
25-29	62	3.250	3.250	2.667	2.375	2.333	3.000
30-34	63	3.250	3.500	2.333	2.250	1.667	3.000
35-39	59	3.250	3.500	2.333	1.750	1.667	3.000
40-44	76	3.500	3.500	2.500	2.125	2.000	3.000
45-49	26	3.500	3.625	2.333	1.750	2.000	3.333
50-54	46	3.000	3.250	2.500	2.000	1.667	2.333
55 and above	48	3.000	3.375	2.167	1.500	1.333	2.333

Gender		U = 25727.0 p = 0.456	U = 25538.0 p = 0.389	U = 24320.5 p = 0.087	U = 23486.0 p = 0.021*	U = 22602.0 p = 0.003*	U = 23263.0 p = 0.014*
Female	210	3.250	3.250	2.333	1.500	1.667	2.667
Male	255	3.250	3.500	2.333	2.000	2.000	3.000
Race		$\varkappa^{2(2)}$ = 1.6642 p = 0.435	$\varkappa^{2(2)}$ = 0.045 p = 0.978	$\varkappa^{2(2)}$ = 3.751 p = 0.153	$\varkappa^{2(2)}$ = 1.916 p = 0.384	$\varkappa^{2(2)}$ = 0.084 p = 0.959	$\varkappa^{2(2)}$ = 2.140 p = 0.343
Chinese	217	3.250	3.250	2.333	1.750	1.667	2.667
Indian	59	3.000	3.500	2.000	1.500	1.667	3.000
Malay	189	3.250	3.500	2.333	2.000	1.667	3.000
Working Industry[#]		$\varkappa^{2(10)}$ = 18.343 p = 0.049*	$\varkappa^{2(10)}$ = 10.808 p = 0.373	$\varkappa^{2(10)}$ = 20.271 p = 0.027*	$\varkappa^{2(10)}$ = 24.298 p = 0.007*	$\varkappa^{2(10)}$ = 8.279 p = 0.602	$\varkappa^{2(10)}$ = 20.936 p = 0.022*
WI1	32	2.750	3.250	2.333	1.750	1.667	2.833
WI2	25	3.500	3.500	2.333	2.000	2.000	3.000
WI3	38	2.750	3.000	1.667	1.250	1.667	2.333
WI4	26	3.250	3.500	2.000	1.500	1.500	2.833
WI5	39	3.500	3.500	2.000	2.000	1.667	3.000
WI6	36	3.000	3.000	2.667	2.125	2.000	2.667
WI7	10	3.750	3.750	2.833	2.250	2.000	3.167
WI8	13	4.250	4.000	2.333	2.000	1.667	3.667
WI9	23	3.250	3.250	2.333	1.250	1.667	2.000
WI10	75	3.500	3.250	2.333	2.250	1.667	3.000
WI11	148	3.000	3.500	2.667	1.875	1.667	3.000

* Median comparison is significantly different

[#] WI1 = Education; WI2 = Health / Insurance; WI3 = Student; WI4 = Banking and financial institution; WI5 = Direct Selling / Retailer; WI6 = Audit / Accountancy / Legal; WI7 = Tourism and Hospitality; WI8 = Government agencies; WI9 = Telecommunication; WI10 = Architecture / Engineering / Real estate / Transportation / Utilities / Wholesale; WI11 = Others

For gender groups, as observed in Table 9, there were significant differences ($p < 0.05$) between males and females in disclosing personal data categories of Finance, Authenticating and Medical-Health. The results showed that comparatively females were less willing to disclose these three categories of personal information. On the other hand, age group 55 and above was found the least likely to disclose Authenticating and Medical-health information. Students scored the lowest median score indicating the least likely to disclose Tracking and Finance information, followed by age group 55 and above.

There was divergence in the respondents' disclosure intention across industries ($p < 0.05$), particularly associated with data categories of Tracking, Finance and Medical-health. For the Tracking data category, students were found the least likely group to disclose, followed by Banking and Financial institution, and Direct Selling/Retailer sectors. Students and Telecommunication sector were the least willing to disclose Financial information. Conversely, respondents from Government agencies scored the highest median score among sectors in disclosing Medical-health information.

There was no statistically significant difference between races in disclosing different data categories.

Table 10. Demographics Differences between Personal Data Categories for Perceived Privacy Concern

	N	**Lifestyle-behaviour** **Test/p-value Median**	**Social-economic** **Test/p-value Median**	**Tracking** **Test/p-value Median**	**Finance** **Test/p-value Median**	**Authenticating** **Test/p-value Median**	**Medical-health** **Test/p-value Median**
Age		$\varkappa^{2(7)}$ = 6.988 p = 0.430	$\varkappa^{2(7)}$ = 3.203 p = 0.866	$\varkappa^{2(7)}$ = 15.501 p = 0.030*	$\varkappa^{2(7)}$ = 12.751 p = 0.078	$\varkappa^{2(7)}$ = 11.137 p = 0.133	$\varkappa^{2(7)}$ = 11.174 p = 0.131
Below 25	85	3.250	3.250	4.333	4.750	4.667	3.333
25-29	62	3.250	3.125	4.000	4.500	4.333	3.667
30-34	63	3.250	3.250	4.333	4.250	4.667	3.667
35-39	59	3.500	3.500	4.000	4.250	4.000	3.667
40-44	76	3.500	3.000	4.000	4.250	4.667	3.667
45-49	26	3.125	3.250	4.000	4.750	4.833	4.000
50-54	46	3.500	3.250	4.000	4.000	4.667	3.667
55 and above	48	3.000	3.000	3.667	4.250	4.333	3.167

Gender		U = 25190.0 p = 0.270	U = 25155.5 p = 0.260	U = 25490.0 P = 0.360	U = 26310.0 p = 0.742	U = 25537.5 p = 0.377	U = 26541.0 p = 0.870
Female	210	3.250	3.250	4.000	4.500	4.667	3.667
Male	255	3.000	3.250	4.000	4.500	4.667	3.667
Race		$\varkappa^{2(2)}$ =1.591 p = 0.451	$\varkappa^{2(2)}$ = 1.475 p = 0.478	$\varkappa^{2(2)}$ = 0.406 p = 0.816	$\varkappa^{2(2)}$ = 1.186 p = 0.553	$\varkappa^{2(2)}$ = 0.178 p = 0.915	$\varkappa^{2(2)}$ = 0.371 p = 0.831
Chinese	217	3.250	3.250	4.000	4.500	4.667	3.667
Indian	59	3.250	3.500	4.000	4.500	4.667	3.667
Malay	189	3.250	3.250	4.000	4.500	4.333	3.667
Working Industry#		$\varkappa^{2(10)}$ = 11.563 p = 0.315	$\varkappa^{2(10)}$ = 10.808 p = 0.787	$\varkappa^{2(10)}$ = 20.271 p = 0.087	$\varkappa^{2(10)}$ = 24.298 p = 0.180	$\varkappa^{2(10)}$ = 8.279 p = 0.821	$\varkappa^{2(10)}$ = 20.936 p = 0.800
WI1	32	3.375	3.375	4.333	4.500	4.667	3.667
WI2	25	3.500	3.250	4.667	5.000	4.667	3.667
WI3	38	3.250	3.250	4.500	4.750	4.667	3.167
WI4	26	2.750	3.000	4.333	4.500	4.500	3.667
WI5	39	3.250	3.000	4.333	4.500	4.667	3.667
WI6	36	3.750	3.750	4.333	4.500	4.667	3.667
WI 7	10	2.625	3.250	3.333	4.375	4.667	3.167
WI 8	13	3.750	3.500	3.667	4.250	4.333	3.333
WI 9	23	3.250	3.250	4.333	4.750	4.667	3.333
WI 10	75	3.500	3.250	4.000	4.250	4.333	3.667
WI 11	148	3.000	3.000	4.000	4.250	4.500	3.667

* Median comparison is significantly different

[#] WI1 = Education; WI 2 = Health / Insurance; WI 3 = Student; WI 4 = Banking and financial institution; WI5 = Direct Selling / Retailer; WI6 = Audit / Accountancy / Legal; WI7 = Tourism and Hospitality; WI8 = Government agencies; WI9 = Telecommunication; WI10 = Architecture / Engineering / Real estate / Transportation / Utilities / Wholesale; WI11 = Others

For perceived privacy concern, there was no statistically significant difference in perceived privacy concern among races, gender and working industry. Exceptionally, for the Tracking data category, age distribution among groups showed significant differences ($p < 0.05$), with the tendency being the younger the age groups, the higher their privacy concern score.

5. Discussion

5.1 Main Findings

Our research outcomes present a validated finding in personal data categorization. The results of inter-coding tests via Cohen's κ and factor analysis confirmed the validity and reliability of the data categories associating with the characteristics we identified in this study. The findings also proved that different personal data categories have significantly different levels of perceived disclosure intention and information privacy concern. As diagrammatically presented in Figure 1, overall perceived information privacy concern showed an opposite tendency compared to disclosure intention, with the exception of the Tracking data category, which presented the opposite phenomenon between the two mean ranks.

Although our respondents significantly showed concern with regards to the Tracking category information, contradictorily they were found likely to disclose this information nevertheless. This result reflects individuals' conflicting attitude associated with Tracking information. In real life, disclosing Tracking information is required to enable service provision or communication. For example, contact information is needed to communicate with others or allows service providers to contact individuals, whereas location-based information is necessary to enable navigation service. This finding provides an extended view of the privacy-paradox attitude from the dimension of data categorization, in that individuals' concern with Tracking category information is more likely to be overridden by the desire of using an application, given gratification or time constraints (Barth & De Jong, 2017) compared to other data categories.

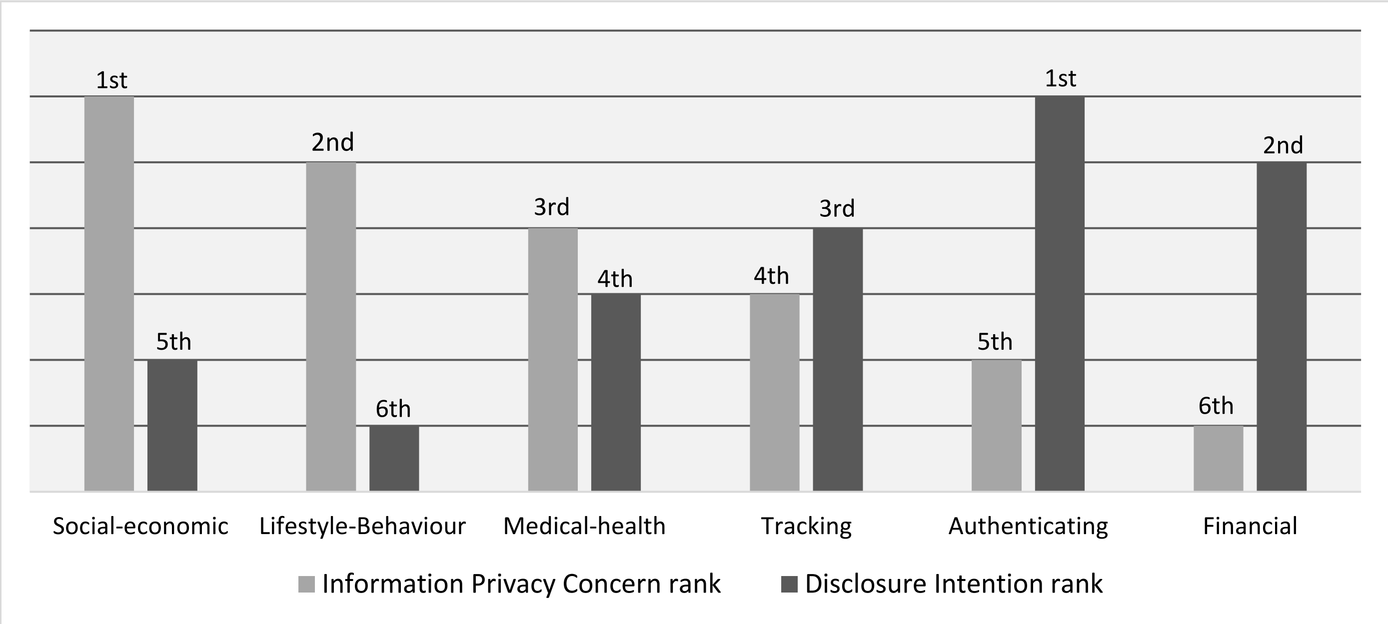

Figure 1. Rank Comparisons between Information Privacy Concern and Disclosure Intention

Besides that, it was found that the Authenticating and Financial categories of personal data posed the highest level of privacy concern compared to other categories while having the lowest level of disclosure intention. The Authenticating and Finance categories shared a comparable level of high privacy concern and low disclosure intention. In rationale, if the information of Authenticating category was exposed and misused, personal data from the Financial category could be potentially obtained as well, through confidential account login information as an example.

With today's common use of online social media platforms, individuals could easily share their daily activities anytime anywhere by posting their life stories, thoughts and opinions towards incidents or events that may expose their social demographic information and lifestyle as well as their personal behavioral characteristics. Our results statistically confirm this phenomenon through the observation of the Social-economic and Lifestyle-behavior categories, which demonstrated comparatively lowest level of privacy concern, and hence, not surprisingly, the most likely categories of personal information for disclosure intention. For Medical-health information, our respondents showed moderate concern and were likely to disclose this information reasonably. This could be because individuals are usually required to report on their medical history and health diagnosis results or conditions prior to getting treatments.

Although the Tracking and Medical-health categories do not have similar characteristics in nature, they share a common 'purpose', that is, the information is required to achieve something that an individual want, such as treatment, service or communication. Regardless of this common purpose, we argue that these two categories should be treated as separate categories because they are proven significantly different in both information privacy concern and disclosure intention. Furthermore, individuals' concern of being tracked was higher than their concern regarding their medical information being exposed (as shown in Table 5), leading to a greater willingness to disclose their medical information (as shown in Table 7). The mean rank and pairwise comparison results were in line with the findings of prior studies (Phelps et al., 2000; Anderson & Agarwal, 2011; Bansal & Gefen, 2010; Jersey & Chua, 2018) which showed that individuals were generally unlikely to disclose their personal data if they had greater privacy concerns.

Our study discovered some noteworthy results regarding the effect of demographic factors on perceived information privacy concern and disclosure intention for the different data categories. The effects of gender and age are important variables to consider.

Females were found less willing to disclose personal information especially related to more confidential data categories such as Finance, Authenticating, and Medical-Health. This observation may be explained by integrating the findings of Dutton and Shepherd (2006) indicating that the higher computer proficiency, the less likely an individual be concerned with associated risks that lead to more willingness to disclose information, and Zin et al. (2000) showing Malaysian females have lower computer literacy compared to males among undergraduate students.

Our findings indicate that younger individuals are more concerned and less likely to disclose Tracking and Finance information compared to other age groups. This finding contradicts Prensky's (2001) study that shows younger individuals as "digital natives" who have grown up and feel comfortable with technology access demonstrate a more positive attitude toward disclosing personal data. The contradiction may be explained with the rationale that younger individuals are more proficient in internet technology use, and therefore have more awareness of potential threats of computer hacking that may lead to Financial loss or the awareness of technology's capability in using their computer device details for tracking their online activities, for example, data collection by online service platforms such as Facebook or Google allows the service providers knowing websites one visits or one's social/political connections. This awareness was found positively associated with privacy concerns (Raider, 2014), and negatively influence the likelihood of disclosing personal information (Nemec Zlatolas et al., 2015).

Inconsistent with previous literature that shows older group individuals are more likely to be concerned about privacy (Van den Broeck et al., 2015; Kazer et al., 2016), our findings reveal a conflicting observation that the older age group (>55) has no significant privacy concern tendency compared to most of the age groups (except for Tracking information), however, they are least likely to disclose Authenticating and Medical-health information. One possible explanation to be considered is probably that privacy concern of Malaysian older age group does not factor into their disclosure decision when involving Authenticating and Medical-health information. While many prior studies (Joinson et al., 2010; Lo, 2010; Nemec Zlatolas et al., 2015) show a significant association between privacy concerns and disclosure intention, other literature fails to associate individuals' privacy concerns with their disclosure behaviors (Taddicken, 2014). Our observation on the older age group suggests that with the transparency of different personal data categories, privacy concerns might not always be the factor associating with disclosure intention. This suggestion infers willingness to disclose information considers both concern and specific disclosure categories of personal data.

In addition to prior studies' contribution however, this study extends the understanding of information privacy concern and disclosure intention by providing a more fine-grained insight of how they shift when associating with different personal data categories. Table 11 presents a summarized comparison between our present research and prior studies related to personal data categorization.

Table 11. Comparisons between prior studies and present research

Research Findings	**Present Research**	**Phelps et al. (2000)**	**Robinson (2016)**	**Milne et al. (2017)**	**Park et al. (2018)**	**Rumbold & Pierciokneк (2018)**
Mechanism used to form personal data categories	Aggregation based on results of prior studies	*(Not mentioned)* *Note: Structured according to the nature of data characteristics)*	*(Not mentioned)*	Clustering method	*(Not mentioned)* *Note: Structured based on the nature of data characteristics*	*Not mentioned)* *Note: Structured based on sensitivity and nature of data characteristics*
Validity of personal data categorization	Validated with Cohen's κ test, Cronbach's Alpha (α), Average Variance Extracted (AVE), and Composite Reliability (CR)	*(Not mentioned)*	*(Not mentioned)*	Validated with clusters' F- to compare the variability between data categories' means	*(Not mentioned)*	*(Not mentioned)*
Dimensions of differences between personal data categories	Perceived privacy concern and disclosure intention	*(Not mentioned)*	*(Not mentioned)* Investigated the impact of personal identifiable information (PII) as a whole instead of different categories on perceived risk and disclosure intention)	Perceived risk, disclosure and sensitivity by customer segments	Perceived value priority of personal information type	*(Not mentioned)*
Validity of differences between personal data categories	Validated with Friedman and Wilcoxon tests	*(Not mentioned)*	*(Not mentioned)*	*(Not mentioned)* *Note: Validity test on customer segments level instead of personal data categorization level*	*(Not mentioned)* *Note: Ranked different information types instead of personal data categories*	*(Not mentioned)*
Significant influence of	Disclosure intention:	*(Not mentioned)*	*(Not mentioned)*	*(Not mentioned)*	*(Not mentioned)*	*(Not mentioned)*

demographic factors on perception associated with personal data categorization	• Age, Gender, Working Industry Privacy concern: • Age, Gender, Working Industry		Investigated the impact of demographic factors on perceived risk and disclosure intention without associating different personal data categories	*Note: Analysed demographics influence associated with customer segments instead of personal data categories*		
Country of respondents (sample size)	Malaysia (465)	America (555)	America (257), Estonia (297)	America (310)	Korea (44)	*(No data collection mentioned)*

Our study provides new evidence regarding validated personal data categories and their significant differences in perceived information privacy concern and disclosure intention. Our research findings also discovered that Age, Gender and Working Industry as demographic factors had significant effects on the disclosure intention associated with Tracking, Finance, Authenticating and Medical-health information.

5.2 Contributions and Implications

The core contribution of this study is our validated personal data categorization, and the novel finding that different personal data categories are perceived significantly different in relation to information privacy concern and disclosure intention. With the evidence presented in this study, i.e. different categories of personal data correlate with different levels of concern and disclosure intention, this research provides a more in-depth view on personal data, demonstrating that personal data should not be treated as a singular category. By referring to the validated personal data categorization as a guideline, our research outcomes bring implications to several stakeholders in their personal data protection strategy and implementation.

Implications for lawmakers:

- Our finding of personal data categorization enables a clearer differentiation of personal data categories, consequently avoiding service providers' requests for loose permissions on personal data including sensitive information that might be irrelevant and unnecessary for the use of the provided services. Consequently, this enables the demand of a finer requirement on service providers for stricter permissions on different personal data categories that are only relevant and necessary for the use of the provided services.
- Besides, authorities would be able to differentiate the amount of fine or the extent of enforcement measures posed on the misuse of different categories of personal data, as the categorization provides an understanding of the differences between categories of personal data based on their perceived importance levels. With this categorization, authorities could impose a fairer punishment depending on the different data categories involved, and the relevant stakeholders would be informed of the severity of the problem which could ultimately lead to assessments of the appropriate level of protection needed on different categories. For instance, the leaking of data related to Finance and Authenticating categories, which should require a top level of protection, would be imposed a higher level of punishment compared to other data categories.
- We also urge the authorities to conduct further research to capture an understanding of individuals' opinion of their conflicting privacy-paradox attitude as well as how certain data categories such as Tracking information can be better protected through regulation enforcement in order to decrease concern.

Implications for organizations:

- With the understating of different perceived important levels of data categories, organizations would be able to conduct more category-specific evaluation in their data processing to enhance the level of security on each personal data category. Access control to different categories should be imposed with different restriction levels.
- Our findings indicate a requirement for system designers and developers to consider a personal data category-specific approach in modelling user personal profile, identity management and data access control mechanisms.
- Organizations can better formulate their communications with their customers with this understanding of the different levels of privacy concern and disclosure intention associated with different personal data categories.
- In addition, organization privacy policies could also reflect this understanding in a more nuanced manner, by taking into consideration the differing privacy concerns associated with different personal data categories.

Implications for individuals as consumers:

- Personal data categorization enables greater transparency for individuals as service users/customers in terms of understanding what category of their personal data is relevant for the service provided; this could allow them to exercise their right to choose not to disclose irrelevant data categories instead of being forced to provide unnecessary data, as is likely the case when personal data is treated as a singular category.

- Regulations imposed on digital content provided in exchange of personal data indicate the financial importance of individuals' personal data. The identification and awareness of personal data categorization could allow individuals to demand a better monetary offer and protection based on different personal data categories. This swift the power of individuals from being passively forced to disclose not only relevant but also irrelevant personal data unnecessarily as a singular category for use of services.

Implications for the research community:

- Our initial work can be a foundation for future research to build upon. Different demographics from other countries and samples with additional factors could be tested as perceived privacy and disclosure intention are contextually driven (Chua et al., 2018; Sheehan, 1999; Albrechtsen, 2007),
- The concept of Privacy by design (Cavoukian, 2009) calls for privacy to be considered throughout the whole system engineering process. This concept takes human values into account in a well-defined manner throughout the whole process. The different personal data categories with different levels of concern and disclosure intention put forth a design guideline for modelling user identity and management, something which needs to be taken into consideration at the beginning of a system design. This is because user identity modelling and management aspects can be shaped by their personal data characteristics, which eventually influence the database structure and data relationship especially with the type of services provided, security levels, and access control mechanisms.

5.3 Limitations and Future Research

This study comprises some limitations which would require additional research. Firstly, we were only able to collect a sample size of 465 respondents in Malaysia, which may not be representative enough to enable us to generalize the results to the Malaysian population due to lack of data showing Malaysian demographic information from age 18 and above. Further, our study might not reflect similar results in the research of respondents' perception from other countries. Therefore, in order to expand the generalization to populations of other countries, more responses would need to be collected in the future.

Future research work extending this study will be investigating the mechanisms and challenges of incorporating personal data categorization into user identify management, and how the implementation of these mechanisms impacts the whole system engineering process and user interfaces.

5.4. Conclusion

To conclude, our research questions have been answered and the study has confirmed that different categories of personal data indeed have significant differences in terms of perceived information privacy concern and disclosure intention. Our research study identified and validated six distinct personal data categories: Social-economic, Lifestyle-behavior, Tracking, Financial, Authenticating, and Medical-health. Organizations can use these validated personal data categories to provide more transparency in how each personal data category will be processed and used. This transparency could build an individual's confidence and trust towards an organization. Besides, our study can help regulators to recognize different personal data categories to formulate a standard for measurement in realizing the requirement of "the processed personal data must be adequate, relevant and limited to what is necessary for the purposes for which it is processed". The terms "adequate", "relevant" and "necessary" can now be more measurable with the understanding of different personal data categories, and the magnitude of different categories' impact on individuals' concern over the potential threat of disclosing the data. Our findings provide new insights by offering a more fine-grained understanding of personal data for better data protection through category-specific system design, stricter regulatory requirements, and more transparency in data collection. Our study to identify the effects of demographic factors leads to original evidence that implies disclosure behavior of different age groups and gender take into account both privacy concern and specific disclosure categories of personal data.

Acknowledgment:
Funding: This research was supported by the Malaysian government FRGS grant [FRGS/1/2019/ICT04/SYUC/02/2].

References:

Acquisti, A., Taylor, C., & Wagman, L. (2016). The economics of privacy. Journal of Economic Literature, 54(2), 442-92.

Ahern, S., Eckles, D., Good, N. S., King, S., Naaman, M., & Nair, R. (2007). Photo sharing: Over-exposed?: Privacy patterns and considerations in online and mobile photo sharing. Proceedings of CHI '07. New York: ACM.

Albrechtsen, E., 2007. A qualitative study of users' view on information security. Comput. Secur. 26, 276–289.

Anderson, C.L. and Agarwal, R., 2011. The digitization of healthcare: boundary risks, emotion, and consumer willingness to disclose personal health information. Information Systems Research, 22(3), pp.469-490.

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(2), pp.138-150.

Barth, S., & De Jong, M. D. (2017). The privacy paradox–Investigating discrepancies between expressed privacy concerns and actual online behaviour–A systematic literature review. Telematics and Informatics, 34(7), 1038-1058.

Bellotti, V., & Sellen, A. (1993). Design for privacy in ubiquitous computing environments. Proceedings of the Third European Conference on Computer-Supported Cooperative Work 1993, 77–92. Norwell, MA: Kluwer Academic Publishers.

Bertea, P., & Zait, A. (2011). Methods for testing discriminant validity. *Management Marketing Journal*, *9*(2), 217-224.

Brannon, J. B., Jones, K., Patton-Kuhl, D. D., Kveen, B. P., Pavlichek, N. I., Crawford, E. R., ... & Shah, M. (2020). U.S. Patent No. 10,614,247. Washington, DC: U.S. Patent and Trademark Office.

Brunner, M. & Süβ, H.M., 2005. Analyzing the reliability of multidimensional measures: An example from intelligence research. *Educational and Psychological Measurement*, *65*(2), pp.227-240.

Cavoukian, A., 2009. Privacy by design: The 7 foundational principles. Information and privacy commissioner of Ontario, Canada, 5.

Chang, Y., Wong, S. F., Libaque-Saenz, C. F., & Lee, H. (2018). The role of privacy policy on consumers' perceived privacy. Government Information Quarterly, 35(3), 445–459.

Chua, H., Wong, S., Low, Y., & Chang, Y. (2018). Impact of employees' demographic characteristics on the awareness and compliance of information security policy in organizations. Telematics and Informatics, 35(6), 1770-1780.

Conover, W. J. (1998). Practical nonparametric statistics (Vol. 350). John Wiley & Sons.

Cradock, E., Stalla-Bourdillon, S., & Millard, D. (2017). Nobody puts data in a corner? Why a new approach to categorising personal data is required for the obligation to inform. Computer Law and Security Review, 33(2), 142–158.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of a test. *Psychometrika*, *16*, 297-334.

Culnan, M., P. Armstrong. 1999. Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation. Organ. Sci. 10(1) 104–115.

Culnan, M., R. J. Bies. 2003. Consumer privacy: Balancing economic and justice considerations. J. Soc. Issues 59(2) 323–342.

Degele, J., Hain, J., Kinitzki, V., Krauß, S., Kühfuß, P., & Sigle, N. (2017). Data architecture for digital health insurances. Digital Enterprise Computing.

Derrick, B., & White, P. (2017). Comparing two samples from an individual Likert question. International Journal of Mathematics and Statistics, 18(3).

Dinero, J., & Chua, H. N. (2018, November). Predicting Personal Mobility Data Disclosure. In 2018 IEEE Conference on Big Data and Analytics (ICBDA) (pp. 1-6). IEEE.

Dinev, T., P. Hart. 2006. An extended privacy calculus model for e-Commerce transactions. Inform. Systems Res. 17(1) 61–80.

Drinkwater, D. (2016). Does a data breach really affect your firm's reputation. Online source from http://www.csoonline.com/article/3019283/data-breach/does-a-data-breach-really- affect-your-firm-s-reputation.html. Last accessed on 15th April 2019.

Dutton, W.H., Shepherd, A., 2006. Trust in the Internet as an experience technology. Inform. Commun. Soc. 9 (4), 433–451. http://dx.doi.org/10.1080/13691180600858606.

Dwyer, C., Hiltz, S.R., & Passerini, K. (2007). Trust and Privacy Concern Within Social Networking Sites: A Comparison of Facebook and MySpace. AMCIS.

Enck, W., Gilbert, P., Han, S., Tendulkar, V., Chun, B. G., Cox, L. P., & Sheth, A. N. (2014). TaintDroid: an information-flow tracking system for realtime privacy monitoring on smartphones. ACM Transactions on Computer Systems (TOCS), 32(2), 5.

Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big data consumer analytics and the transformation of marketing. Journal of Business Research, 69(2), 897–904.

GDPR (2018). European Union General Data Protection Regulation, 2018. Online source from https://eur-lex.europa.eu/eli/reg/2016/679/oj. Last accessed on 15th April 2019.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2009). Multivariate Data Analysis (7th ed.). Upper Saddle River, New Jersey: Pearson Education Limited.

Hunn, E. (1979). Cognition and Categorization. Eleanor Rosch, Barbara B. Lloyd. American Anthropologist, 81(3), 712-713. DOI: 10.1525/aa.1979.81.3.02a00710.

Iapp. (n.d.). Categories of Personal Data. Retrieved June 02, 2018, source from https://iapp.org/resources/article/categories-of-personal-data/. Accessed on 12th April 2019.

Janssen, M., & Kuk, G. (2016). The challenges and limits of big data algorithms in technocratic governance. Government Information Quarterly, 33(3), 371–377

Janssen, M., & van den Hoven, J. (2015). Big and open linked data (BOLD) in government: A challenge to transparency and privacy? Government Information Quarterly, 32(4), 363–368.

Joinson, A. N. (2008). “Looking at,” “Looking up” or “Keeping up with” people? Motives and uses of Facebook. Proceedings of CHI 2008 New York: ACM.

Joinson, A. N., Reips, U. D., Buchanan, T., & Schofield, C. B. P. (2010). Privacy, trust, and self-disclosure online. Human–Computer Interaction, 25(1), 1-24.

Kezer, M., Sevi, B., Cemalcilar, Z., & Baruh, L. (2016). Age differences in privacy attitudes, literacy and privacy management on Facebook. Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 10(1).

Li, H., Sarathy, R., & Xu, H. (2011). The role of affect and cognition on online consumers’ decision to disclose personal information to unfamiliar online vendors. Decision Support Systems, 51(3), 434–445.

Lo, J. (2010, August). Privacy Concern, Locus of Control, and Salience in a Trust-Risk Model of Information Disclosure on Social Networking Sites. In AMCIS (p. 110).

Malgieri, G., & Custers, B. (2017). Pricing privacy - the right to know the value of your personal data. Computer Law and Security Review. DOI: https://doi.org/10.1016/j.clsr.2017.08.006.

Malhotra, N. K., S. S. Kim, J. Agarwal. 2004. Internet users’ information privacy concerns (IUIPC): The construct, the scale, and a causal model. Inform. Systems Res. 15(4) 336–355.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. Biochemia medica: Biochemia medica, 22(3), 276-282.

McKnight, D. H., V. Choudhury, V. C. Kacmar. 2002. Developing and validating trust measures for e-Commerce: An integrative topology. Inform. Systems Res. 13(3) 334-359.

McKnight, P. E., & Najab, J. (2010). Kruskal-wallis test. The corsini encyclopedia of psychology, 1-1.

McKnight, P. E., & Najab, J. (2010a). Mann-Whitney U Test. The Corsini encyclopedia of psychology, 1-1.

Milne, G. R., Pettinico, G., Hajjat, F. M., & Markos, E. (2017). Information sensitivity typology: Mapping the degree and type of risk consumers perceive in personal data sharing. Journal of Consumer Affairs, 51(1), 133-161.

Morey, T., Forbath, T., & Schoop, A. (2015). Customer data: Designing for transparency and trust. Harvard Business Review, 93(5), 96–105.

Muffat, C., & Kodliuk, T. (2020). U.S. Patent Application No. 16/731,351.

Nemec Zlatolas, L., Welzer, T., Heričko, M., & Hölbl, M. (2015). Privacy antecedents for SNS self-disclosure. Computers in Human Behavior, 45(C), 158-167.

Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The privacy paradox: Personal information disclosure intentions versus behaviours. Journal of consumer affairs, 41(1), 100-126.

Park, M., Chai, S., Azyabi, N. G., Lou, L., Koh, J., Park, J., & Rho, H. (2018). The Value of Personal Information: An Exploratory Study for Types of Personal Information and Its Value. Asia Pacific Journal of Information Systems, 28(3), 154-166..

Petronio, S. 2002. Boundaries of Privacy: Dialectics of Disclosure. SUNY Press, Albany, NY.

PDPA (2013). The Malaysian Personal Data Protection Act. 2013. Online source from http://www.agc.gov.my/agcportal/uploads/files/Publications/LOM/EN/Act%20709%2014%206%202016.pdf. Last accessed on the 15th July 2019.

Phelps, J., Nowak, G. & Ferrell, E., 2000. Privacy concerns and consumer willingness to provide personal information. Journal of public policy & marketing, 19(1), pp.27-41.

Rader, E. (2014). Awareness of behavioral tracking and information privacy concern in facebook and google. In 10th Symposium On Usable Privacy and Security ({SOUPS} 2014) (pp. 51-67).

Robinson, C. (2017). Disclosure of personal data in ecommerce: A cross-national comparison of Estonia and the United States. Telematics and Informatics, 34(2), 569-582.

Rumbold, J. M. M., & Pierscionek, B. K. (2018). What Are Data? A Categorization of the Data Sensitivity Spectrum. Big Data Research, 12(November), 49–59. DOI: https://doi.org/10.1016/j.bdr.2017.11.001

Shapiro, S.S. and Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3/4), pp.591-611.

Sheehan, K.B., 1999. An investigation of gender differences in on-line privacy concerns and resultant behaviors. J. Interact. Mark. 13, 24–38.

Sidgman, J., & Crompton, M. (2016). Valuing personal data to foster privacy: A thought experiment and opportunities for research. Journal of Information Systems, 30(2), 169-181..

Smith, H.J., Milberg, S.J. and Burke, S.J., 1996. Information privacy: measuring individuals' concerns about organizational practices. MIS quarterly, pp.167-196.

Soanes, C. (2011). Oxford English mini dictionary. New York: Oxford.

Son, J.Y. & Kim, S.S., 2008. Internet users' information privacy-protective responses: A taxonomy and a nomological model. MIS quarterly, pp.503-529.

Statista. (2019). Number of mobile phone users worldwide from 2015 to 2020 (in billions). Online source from https://www.statista.com/statistics/274774/forecast-of-mobile-phone-users-worldwide/. Last accessed on the 6th August 2019.

Swisher, L. L., Beckstead, J. W., & Bebeau, M. J. (2004). Factor Analysis as a Tool for Survey Analysis Using a Professional Role Orientation Inventory as an Example. *Physical Therapy*, *84*(9), 784-799. doi: 10.1093/ptj/84.9.784.

Taddicken, M. (2014). The ‘privacy paradox’ in the social web: The impact of privacy concerns, individual characteristics, and the perceived social relevance on different forms of self-disclosure. Journal of Computer-Mediated Communication, 19(2), 248-273.

TRUSTe. (2011). Smart Privacy for Smartphones: Understanding and delivering the protection consumers want. Online resource from www.truste.com. Last accessed on the 6th August 2019.

Tsai, J.Y., Egelman, S., Cranor, L., & Acquisti, A. (2011). The effect of online privacy information on purchasing behaviour: an experimental study. Inf. Syst. Res. 22, 254–268.

Van den Broeck, E., Poels, K., & Walrave, M. (2015). Older and wiser? Facebook use, privacy concern, and privacy protection in the life stages of emerging, young, and middle adulthood. Social Media+ Society, 1(2). http://dx.doi.org/10.1177/2056305115616149.

Vroom, V. H. 1964 Work and Motivation. Wiley, New York.

Wang, X., & Peng, X. (2013). Research on data leak protection technology based on TPM. Proceedings - 2013 International Conference on Mechatronic Sciences, Electric Engineering and Computer, MEC 2013, 2354–2358.

Yap, B. W. & Sim, C. H. (2011). Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation, 81*(12), 2141-2155. DOI: 10.1080/00949655.2010.520163

Zhang, B., & Zhang, Y. (2009). Mann-Whitney U test and Kruskal-Wallis test should be used for comparisons of differences in medians, not means: comment on the article by van der Helm-van Mil et al. Arthritis and rheumatism, 60(5), 1565.

Zin, N. A. M., Zaman, H. B., Judi, H. M., Mukti, N. A., Amin, H. M., Sahran, S., ... & Abdullah, Z. (2000). Gender differences in computer literacy level among undergraduate students in Universiti Kebangsaan Malaysia (UKM). The Electronic Journal of Information Systems in Developing Countries, 1(1), 1-8.

APPENDIX: Questionnaire

Section 1: Demographic

Age:	Gender:	Race:	Occupation (industry):

Section 2: How likely are you going to disclose the following data of yours?

	Very unlikely (1)	Unlikely (2)	Neutral (3)	Likely (4)	Very Likely (5)
T1. Contact information (e.g. email address, physical address, telephone number, etc.)					
T2. Communication (e.g. telephone recordings, voice mail, text messages, etc.)					
T3. Location (e.g. country, GPS coordinates, room number, etc.)					
T4. Computer device details (e.g. IP address, Mac address, browser information, digital fingerprint, etc.)					
F1. Credit history (e.g. credit records, credit worthiness, credit standing, credit capacity, etc.)					
F2. Assets (e.g. property, personal belongings, etc.)					
F3. Financial account (e.g. credit card number, bank account, etc.)					
F4. Transactions (e.g. purchases, sales, credit, income, loan records, transactions, taxes, purchases and spending capacity, etc.)					
A1. Passwords or pin (e.g. bank account password or pin, email address password, etc.)					
A2. Identity code (e.g. government issued identification, etc.)					
A3. Username (e.g. social media username, online banking username, etc.)					
MH1. Diagnoses (e.g. drug test results, health records, prescriptions, physical and mental health, disabilities, etc.)					
MH2. Genetic data (e.g. genetic information, blood type, etc.)					
MH3. Personal health history and medication experiences					
LB1. Belief (e.g. religious beliefs, philosophical beliefs, thoughts, etc.)					
LB2. Preferences or interests (e.g. opinions, intentions, interests, favourite foods, colours, likes, dislikes, etc.)					
LB3. Behavior (e.g. browsing habit, call patterns, links clicked, demeanour, attitude, etc.)					
LB4. Relationship (e.g. family structure, siblings, offspring, marriages, divorces, relationships, friends, connections, acquaintances, associations, group membership, etc.)					
SE1. Ethnicity (e.g. race, national / ethnic origin, languages spoken, dialects, accents, etc.)					
SE2. Physical characteristics (e.g. name, picture, etc.)					
SE3. Demographics (e.g. age, gender, etc.)					
SE4. Professional career (e.g. job titles, salary, school attended, employment history, evaluations, references, interviews, certifications, disciplinary actions, know how skills, soft skills, etc.)					

Section 3: How concerned are you towards the following personal data of yours?

	Least concern (1)	Less concern (2)	Neutral (3)	Concern (4)	Most concern (5)
T1. Contact information (e.g. email address, physical address, telephone number, etc.)					
T2. Communication (e.g. telephone recordings, voice mail, text messages, etc.)					
T3. Location (e.g. country, GPS coordinates, room number, etc.)					
T4. Computer device details (e.g. IP address, Mac address, browser information, digital fingerprint, etc.)					
F1. Credit history (e.g. credit records, credit worthiness, credit standing, credit capacity, etc.)					
F2. Assets (e.g. property, personal belongings, etc.)					
F3. Financial account (e.g. credit card number, bank account, etc.)					
F4. Transactions (e.g. purchases, sales, credit, income, loan records, transactions, taxes, purchases and spending capacity, etc.)					
A1. Passwords or pin (e.g. bank account password or pin, email address password, etc.)					
A2. Identity code (e.g. government issued identification, etc.)					
A3. Username (e.g. social media username, online banking username, etc.)					
MH1. Diagnoses (e.g. drug test results, health records, prescriptions, physical and mental health, disabilities, etc.)					
MH2. Genetic data (e.g. genetic information, blood type, etc.)					

MH3. Personal health history and medication experiences					
LB1. Belief (e.g. religious beliefs, philosophical beliefs, thoughts, etc.)					
LB2. Preferences or interests (e.g. opinions, intentions, interests, favourite foods, colours, likes, dislikes, etc.)					
LB3. Behavior (e.g. browsing habit, call patterns, links clicked, demeanour, attitude, etc.)					
LB4. Relationship (e.g. family structure, siblings, offspring, marriages, divorces, relationships, friends, connections, acquaintances, associations, group membership, etc.)					
SE1. Ethnicity (e.g. race, national / ethnic origin, languages spoken, dialects, accents, etc.)					
SE2. Physical characteristics (e.g. name, picture, etc.)					
SE3. Demographics (e.g. age, gender, etc.)					
SE4. Professional career (e.g. job titles, salary, school attended, employment history, evaluations, references, interviews, certifications, disciplinary actions, know how skills, soft skills, etc.)					