ANALYSIS OF PERSONAL DATA-SHARING CONSENT FACTORS, WITH FOCUS ON LOYALTY PROGRAMS IN THE CZECH REPUBLIC

Tomas FORMANEK[1], Radek TAHAL[2]

[1]*Department of Econometrics, Faculty of Informatics and Statistics, University of Economics, Prague, Czech Republic*
[2]*Department of Marketing, Faculty of Business Administration, University of Economics, Prague, Czech Republic*
E-mails: formanek@vse.cz (corresponding author); radek.tahal@vse.cz

Abstract. The purpose of this study is to provide structured, topical and representative analysis of personal data sharing preferences in the Czech Republic. Within the context of personal data sharing and protection, we focus on profiling individuals who voluntarily share their personal data with good-faith corporate entities. Loyalty program operators serve as a common and representative model of commercially driven collection and processing of personal data. We address different types of personal data and factors affecting individual data-sharing consent. Our original research is based on primary surveyed data (806 respondents surveyed during 2017). Multiple quantitative methods such as hierarchical clustering and logistic regression are employed in the analysis. Also, an important part of our research is based on evaluation of structured in-depth interviews, focused on personal data sharing and protection topics. We find pronounced socio-demographic differences in individual propensity to share one's personal data with commercial data processors. Main findings and contrasting factors are pointed out and discussed within the paper. Our analysis reflects the needs of academic and corporate researches to whom it provides actionable and stratified results, especially in context of the new EU legislation: the GDPR directive on personal data protection.

Keywords: personal data, GDPR, data sharing consent, data sharing preferences, socio-demographic factors, loyalty programs, data processing.

JEL Classification: C25, D12, M31.

Introduction

Various aspects of processing and protecting personal data lie at the core of contemporary marketing and business-oriented analyses and discussions, whenever EU's citizens and consumers are involved. The main motivation for such discourse is the new GDPR legislation on the protection of individuals (natural persons) with regard to the processing of their personal data (EUGDPR 2018). Indeed, adjusting for the new GDPR regulation is crucial for many consumer-oriented businesses, their marketing and legal departments and other stakeholders involved. Most medium-sized and large corporate entities are likely to qualify as organizations that engage in large scale systematic monitoring, and/or processing of sensitive personal data (see Article 37 of the GDPR directive). In this respect, we should note that the GDPR legislation distinguishes between data controllers and data processors: while data controllers determine the purpose and means of data usage, data processors are entities that actually work with the data (perhaps on behalf of data controllers).

Today, more than ever before, it is important for marketers to understand how customers perceive sharing of their personal data and/or contact information with business entities, how much confidence or mistrust they feel about it. Under the new legislation, a request for individual consent with data processing must be easily accessible, unambiguous and clearly separated (distinguishable) from other matters communicated with individuals (consumers). At the corporate level, GDPR implementation opens many cost-effectivity questions arising between the complexity and extent of personal data processed and the required consumer's understanding and informed consent to share such

To link to this article: https://doi.org/10.3846/btp.2018.08

extent of data. Ciriani (2015) reflects on this by stating that economic value generated from data processing depends on the regulation of data protection. Some authors highlight the role of CRM activities, where detailed customer data are indispensable. Mitchell (2010) points out that personal information input is crucial for every customer-facing activity: basic data analysis, insight building, innovation through sales and marketing operations, customer services, CRM and even routine administration processes.

Individual willingness to share personal data is determined by many factors. Focusing on data processors, there will be a different degree of trust when data is shared with, say, healthcare providers or government institutions, as opposed to sharing data with corporate entities. The type of personal data requested/shared matters as well: customers can distinguish between necessary data and data that lies beyond the scope of a given interaction. For example: name, postal and e-mail address required by some on-line retailer would seem adequate, while gender and birthdate might be deemed unnecessary (the actual distinction line may vary, given differences in goods or services involved). The propensity to share various types of personal data may be significantly affected by individual's socio-demographic and lifestyle factors (age, gender, education achieved, lifestyle & leisure time preferences, etc.). Overall, the importance of customer data for successful retail operations is widely recognized: for example, Reimer and Becker (2015) emphasize the role of personal data for identification and profiling of profitable customers and their lifetime value.

Based on the above premises, in this analysis we seek three main objectives: (i) we assess and describe the understanding and attitudes of Czech consumers towards personal data protection, their experiences with personal data gathering and sharing; (ii) we aim for a complex analysis of consumer propensity towards sharing different types of personal data with good-faith corporate data processors in the Czech Republic – we use loyalty program participation and the corresponding personal data sharing involved as a frequent and representative model of commercially driven collection and processing of personal data; (iii) we aim to provide relevant and actionable information regarding consumers' data sharing preferences to good-faith corporate entities who seek to find new balance in their data collection and processing operations under the new GDPR legislation.

1. Literature review

Protection of personal data has been increasingly discussed among corporate and academic marketing researchers since the 1990s – ever since information and communication technology became widely available and massively used. For example, Pounder (1998) states: "Organizations which process personal data for their own purposes will have to demonstrate that they have implemented a rigorous

approach with respect to their own IT security arrangements." As a new phenomenon, privacy and data protection topics shift from the real world to the on-line reality where social sites and communication platforms operate. Tavani (1999) points out that some privacy analysts began distinguishing between 'informational privacy' concerns and a separate category related to intrusion and interference that is sometimes referred to as 'psychological privacy'. Recently, Zeman et al. (2017) and other authors also focus on various aspects of military and anti-terrorism cyber security activities.

Around the year 2000, numerous national and international data protection laws and regulatory measures began to emerge in larger numbers. EU's Data Protection Directive 95/46/EC was one of the most influential: "The Directive balances the interests of individuals with the interest of companies that use personal data in their business. The Directive is not designed to ban the collection of data but rather to control its uses." (International Financial Law Review 1998). Interestingly, this very directive is being replaced by the new GDPR legislation. Many authors commenting EU's data protection activities at that time point out and discuss the emphasis on proper manipulation and storage of personal data (Aldhouse 1999). Other authors address issues related to localization of EU's data protection directives to different member countries' legislative environments. Majtenyi (2002) explicitly states that the Czech Republic, Slovakia, Poland and Hungary have made considerable efforts to improve their data protection laws.

The notion of personal data-related risks has grown considerably in recent years. Frederik (2015) focuses on retail sector, where consumer and behavioral targeting activities are often in sharp contrast with personal data protection needs. Kosciejew (2014) notes: "Our personal data is being created, collected, mined, analyzed, monitored, shared, sold, stored, and used for diverse reasons beyond most of our knowledge or control, let alone our willing consent or endorsement." Indeed, the structure and quality of collected customer data has a strategic importance for retailers, their decision-making accuracy and profitability. At the same time, consumers are worried about privacy and fear their personal data could be misused (Tahal et al. 2017). This situation is reflected in numerous other scientific studies – for example, Limba and Šidlauskas (2018) provide an analysis of risks and recommendations aimed at improving the security of personal data of social network site users. Šišulák (2017) examines preventative techniques used against crimes on social networks and analyses their application by competent authorities.

The new GDPR legislation goes a long way in closing the gap between law and the fast-growing personal data-processing industry that has recently outpaced its corresponding legal environment – an issue addressed e.g. by (Kiskis 2010, Tempest and Brandau 2010). At the same time,

the importance of good-faith personal data processing is constantly growing: contemporary consumers increasingly prefer customized (i.e. personalized) forms of marketing communication, which cannot be achieved without appropriate data. Robertshaw and Marr (2006) point out that significant market fragmentation and increased divergence in individual consumer preferences are accelerating the transition from mass-marketing methods to highly personalized marketing programs. Hence, as Cunningham (2012) points out, we are in a process of searching for balance: adequate protective legal framework and individual privacy on one hand and consumers' demand for personalized communication (offers, incentives) on the other hand, along with the need for fair business environment where good-faith marketing activities can be carried out without excessive bureaucratic hurdles. Manzerolle and Smeltzer (2011) present a clear stance on the subject: "Personal information is increasingly the basic fuel on which economic activity runs." Jankalová and Jankal (2017) provide a complex discussion on corporate social responsibility evaluation and corresponding tools, applicable to the field of personal data protection. Limba et al. (2017) examine various types of cyber-security tools for e-government environment and applications.

2. Data and methodology

Our analysis is performed using primary data gathered across the Czech Republic during the year 2017 (March to December). Two complex & representative stratified/quota sampling anonymized surveys (a total of 806 respondents aged 15 and older) were accompanied by a separately organized series of individual in-depth interviews (25 individuals). The surveys were focused on gathering respondents' sociodemographic data, free-time and lifestyle preferences, self-positioning and attitudes toward various types of work and leisure activities, as well as personal data sharing preferences. Different types of questions were used in the survey: quantitative (mostly interval-based), qualitative (Yes/No) and Likert scale (5-degree spans are used). Also, multiple individual in-depth interviews took place in October and November 2017, respondents from the 20÷35 and 36÷50 age groups were approximately equally represented. During each of the 40-minute interviews, three main topics were discussed:

1. Respondents' notion of personal data and their stance on where data protection efforts should be focused.
2. Credibility of different types of entities/authorities that collect and process personal data.
3. What actions respondents undertake to protect their personal data.

The research team (based at the University of Economics, Prague) is led by university employees and teachers who coordinate and supervise the work of students specializing on marketing research. This study is part of a long-term project of systematic surveys and analyses of consumer behavior (Tahal et al. 2017). The surveys used to gather data for this paper may be summarized as follows: Stratified/quota sampling was based on gender, age segmentation and location (domicile) of the respondents. Our empirical analysis is both quantitative (based on logistic regression and related tests and methods) and qualitative (we use output from the in-depth interviews). By combining the quantitative and qualitative approach, we seek to improve interpretability of the results and compare finding based on 'hard-data' against opinions and 'common-sense' stances presented during the interviews.

To carry out the quantitative part of our analysis, the 806-row and 164-column (variable) dataset gathered from the surveys was assembled. Six binary variables describing consumer willingness to share different types of personal information with LP organizers can be ordered by data sensitivity as follows: (a) name & surname, (b) e-mail, (c) address/residence, (d) birthdate, (e) phone number, (f) personal ID number. The first and the last data types (name & surname vs. ID number) have only a limited usage - providing one's name is the very minimum level of engagement a consumer can exert. Besides Czech language-specific gender segmentation, name doesn't carry useful information (as it cannot serve as a unique identificator) and it is usually perceived as the least sensitive data type. Hence, willingness to share and/or deny sharing of name & surname data may serve as a benchmark for propensity to share non-sensitive personal data. At the other end of data sensitivity spectrum, personal ID number is a delicate type of personal data from the GDPR perspective, as it bears a non-negligible misuse potential (along with providing birthdate, gender and unique identification). On the other hand, once data types (b) to (e) are controlled for, personal ID provides little usable information, other than unique identification (which may be addressed otherwise by data processors anyway). Finally, it should be noted that the binary variables indicating willingness/unwillingness to share (a) to (f) types of personal data may be used for two-ways interpretation: we focus on factors influencing the positive response, i.e. consent with sharing one's personal data. Because of the binary nature of the decision on data sharing, our results can be also interpreted in terms of sociodemographic and lifestyle factors reducing individual propensity to share personal data.

From the remaining 158 variables in our survey-generated dataset (leaving out willingness to share different types of personal data), we have empirically selected 120 potential/conceivable explanatory variables describing respondents' sociodemographic status, lifestyle preferences and other relevant information. For detailed quantitative analysis of individual data sharing / data protection preferences, we

combine forward-stepwise selection for logistic models with the non-parametric random forest approach to regressors' importance evaluation (Hastie et al. 2009). Differences in outputs from the two methods are analyzed to detect potential sub optimality in model specification. Our approach allows for assessing the importance of explanatory variables in a way that takes prediction accuracy into account while observing computational feasibility – an exhaustive (brute-force) search for ideal model specification would require 6×2^{120} models to be estimated and evaluated. Although our combined methodology does not guarantee the best model setup possible (at least in theory), it renders empirically well justified model specification (1), with a relatively low potential for sub-optimality.

Evaluation, interpretation and summarizing of the in-depth interviews was performed by senior interviewers. During the interviews, all respondents' answers were recorded in a standardized record sheet, designed and structured along the sections of the interview scenario. Subsequently, the answers gathered were coded and categorized, i.e. respondents' opinions and stances were aggregated into logical content clusters for subsequent evaluation.

3. Empirical results

We begin by verifying some of the basic assumptions that our analysis is built upon, e.g. whether our professional classification of personal data into the (a) to (e) categories is consistent with common customer perception. Without such consistency, we may not be able to correctly interpret the results from our quantitative analyses and draw conclusions with respect to the consumer population. During the in-depth interviews, all individuals mention their name and surname when describing personal data. A large majority of respondents aged 25–35 years would also mention their email & residence addresses among personal data. When asked to elaborate on the notion of personal data, individuals usually focus on internet-based topics: log-in data, contact information provided to online stores (i.e. for ordering and shipping of goods and services), data collected by internet browsers (i.e. cookies, etc.) and other 'hidden' types of on-line collected data that might allow unique personal identification.

Next, respondents of the in-depth interviews were asked about entities to which they would provide personal data without hesitation or fear: most respondents are willing to supply their data to medical doctors, lawyers, municipalities or whenever relevant personal data is obligatory for obtaining requested goods or services (e.g. air travel). By contrast, retail chains and phone operators are listed as entities where respondents are most reluctant to provide personal data. On the other hand, data processors' credibility and renowned brand can play a positive role.

Most of our in-depth interview respondents are convinced that their personal data are handled without adequate diligence and security provisions, once provided to public or private entities. However, the same individual respondents believe that data handling flaws are not in conflict with the legal framework. When asked, almost none of the respondents can depict an actual/presumed breach of terms and conditions of appropriate (legal) handling of personal data. Although such responses might indicate a lack of good understanding of personal data storage and processing operations, we should not jump to quick conclusions here. Even though most respondents have only an intuitive understanding of the subject, they are able to identify different types of potential mishandling of data: in most cases, individuals are highly sensitive to unsolicited telemarketing (phone calls and SMS) and unsought personalized marketing correspondence. Unsolicited printed correspondence is perceived much more negatively than its electronic counterpart (this is in line with our previous findings – see Tahal et al. 2017). After being asked specifically about social media and related personal data protection issues, our respondents usually express data protection concerns related to posting information which becomes 'publicly' available (and a property of the social site operator).

When discussing actual individual experiences with personal data misuse and possible protection measures, about half of the individuals report no previous contact with inappropriate data practices. Those who have such experience usually mention unsolicited telemarking and the reluctance of the callers (telemarketing operators) to disclose the source of their contact data. As for suitable personal data protection measures, respondents suggest to always provide the minimum possible amount of personal data (to operators) and to avoid 'untrustworthy' web pages. About a third of the respondents would autonomously recall the possibility of requesting information from data processors concerning the extent and purpose of storing their personal data. We should note that our in-depth interviews took place at the end of the year 2017, before the EU's GDPR directive activation and it should be stated that the new directive comes a long way in helping individuals with control over their personal data: clear and understandable terms of consent, accountability of data processors and consent revocation terms are changing significantly to improve protection of individual rights and privacy of 'natural persons'.

The aforementioned representative consumer stances and opinions from the in-depth interviews are in line with the methodology of our quantitative analysis. Therefore, the information gathered from in-depth interviews can be reasonably combined with 'hard' surveyed data. Our sample allows us to draw statistically relevant conclusions and the case of sharing personal data for LP participation is a good illustrative example for commercially driven (non-government)
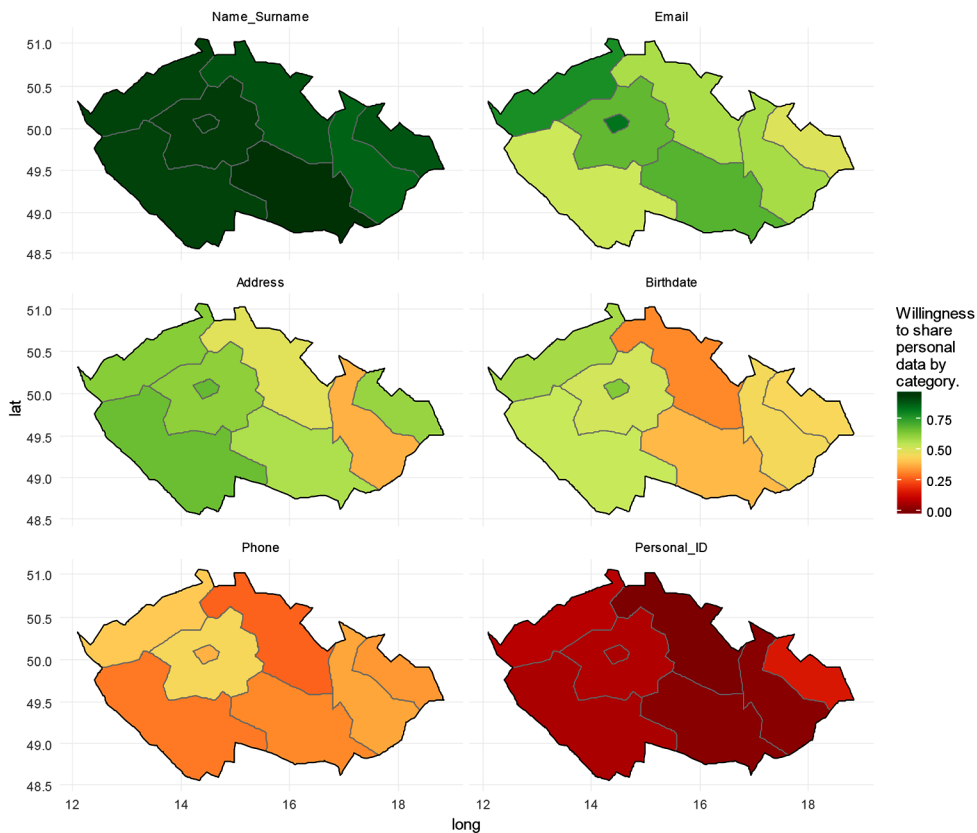
Figure 1. Data sharing preferences across *NUTS2* regions and data categories

personal data collection and processing activities. We begin by summarizing overall consumer propensity to share different categories of personal data: Willingness to share individual's name with a LP organizer (in order to participate in a LP) stands out at 92.8%. Data types (b) and (c) – e-mail and address/residence – have a similar sharing propensity of 67.7% and 60.9%, while willingness to share birthdates with a LP organizer equals 52.1%. Consumer propensity to share phone number is significantly lower – only 32.7% of respondents are willing to provide their phone contact if it's requested for LP participation. This result is in line with our previous analyses (e.g. Tahal et al. 2017) and it also corresponds with the perceived highly intrusive nature of unsolicited telemarketing (declared during the in-depth interviews). The extremely low willingness to share personal ID number with LP organizer – only 6.9% – reflects the delicate nature of this data type and its misuse potential. For reader convenience, the information discussed above is summarized in Figure 1, along with regional variations: at the NUTS2 level, we can see the color-coded differences in personal data sharing preferences. Although most of the differences can be observed between individual choropleths (each 'map' represents one type of personal data), regional differences may be observed as well: respondents in the capital region Prague are slightly more inclined to share personal data as compared to the state-wide average. In

contrast, the North-Eastern region exhibits a consistently lower willingness to share all data types (a) to (f).

The above described average propensities towards sharing personal data provide a solid basic insight. However, we need to bear in mind that similar averages – say, close results for email (b) and address of residence (c) data categories – are not necessarily generated by the same group of respondents willing to share their data (statistically speaking). Theoretically, the two groups of people may be highly divergent – i.e. there might be a very small number of individuals willing to share both (b) and (c) data types simultaneously. From the pairwise correlation analysis, we do not find evidence for any negative relationships among propensities to share different types of data. Typical pairwise correlations for data types (a) to (d) lie between 0.3 and 0.4 (on a 0÷1 scale), while their correlations with the remaining two categories (e) and (f) are significantly lower (yet positive and statistically significant at $\alpha = 5\%$).

Besides pairwise correlations, cluster analysis provides an elegant and interpretable approach to this type of analysis (Murtagh and Legendre 2014). Clustering analysis may be represented in the form of a dendrogram as in Figure 2. Here, we can see the hierarchical nature of the clustering analysis: different personal data categories (leaves of the dendrogram) are clustered into higher level aggregates (branches). For interpretation purposes, an appropriate
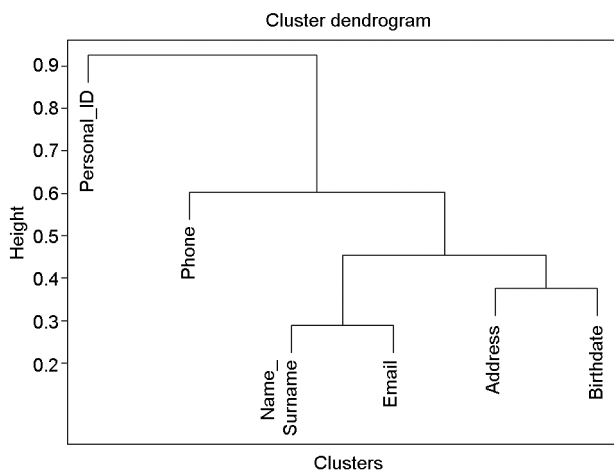
Figure 2. Clustering of personal data sharing preferences

number of clusters is often set arbitrarily. However, considering both the clustering and data type aspects, the dendrogram suggests two main groups of personal data: the first group consists of relatively non-sensitive data and it is formed by data categories (a) to (d); the second category is formed by data types (e) and (f). Although the two types of personal data in the second group (sensitive data) are not particularly homogeneous – i.e. not very similar in terms of individual data sharing propensity – they prominently diverge from the first cluster.

In the final and most detailed step of our analysis, we focus on the socio-demographic and lifestyle determinants of personal data sharing preferences. Using the model-searching algorithm described in previous section, we have established a consistent logistic regression model:

$$y_i = \beta_0 + \beta_1 \, Age.15.24_i + \beta_2 \, Age.25.34_i +$$
$$\beta_3 \, Age.35.49_i + \beta_4 \, Female_i + \beta_5 \, Educ.Univ_i +$$
$$\beta_6 \, LS.exotics.yes_i + \beta_7 \, LS.w.e.house.no_i +$$
$$\beta_8 \, LS.internet.use.yes_i + \beta_9 \, LS.cooking.yes_i +$$
$$\beta_{10} \, LS.shopping.yes_i + \beta_{11} \, LS.paycard.yes_i + R_i + u_i. \, (1)$$

Six separate regression models (1) were estimated, one for each personal data category (a) to (f) taking the place of the dependent variable y_i. Even though all observed regressors in equation (1) are binary, its specification allows for sufficient control over different observed factors so that it allows proper estimation – a situation that is analogous to the Ignorability of treatment assumption (Wooldridge, 2010). Interestingly, surveyed earnings (categorized into 4 levels) came out statistically insignificant and are therefore omitted from equation (1).

Estimation of the logistic regression model (1) was performed using the maximum likelihood approach, which provides consistent and asymptotically normal estimates of the β_j coefficients: $\hat{\beta}_j$. Given the nonlinear nature of logistic regression, individual $\hat{\beta}_j$ coefficient estimates are not particularly informative (except for their sign and statistical

significance). However, convenient and interpretable evaluation of regressors' influence on success probability can be obtained by calculating the average partial effects (APE), as defined in equation (2):

$$APE(x_k) =$$
$$n^{-1} \sum_{i=1}^{n} \left[G\left(\begin{array}{c} \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots \\ + \hat{\beta}_{k-1} x_{i,k-1} + \hat{\beta}_k \end{array} \right) - G\left(\begin{array}{c} \hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \dots \\ + \hat{\beta}_{k-1} x_{i,k-1} \end{array} \right) \right], \, (2)$$

where $APE(x_k)$ determines the average expected effect of the k-th factor on the probability of 'success' (i.e. willingness to share given type of personal data) and $G(.) = \exp(.)/\left[1 + \exp(.)\right]$ is the logistic function. For interpretation of expression (2), we can see that the estimated logistic equation (1) is evaluated twice for each individual: in the first instance, we assume that the k-th factor is present (i.e. $x_{ik} = 1$; please note the $\hat{\beta}_k$ in the first $G(.)$ expression) and then we evaluate the expected probability of 'success' for $x_{ik} = 0$. The differences in probabilities are recorded for each individual and then averaged over the whole sample size. The process is repeated for each of the x_k regressors and the obtained average values $APE(x_k)$, their standard errors (heteroskedasticity in random errors is controlled by using heteroskedasticity-consistent estimators) and statistical significances (p-values) are provided in Table 1.

All six underlying logistic equations (1), estimated for each type of personal data, are statistically significant at the 5% significance level and provide adequate prediction accuracy. Statistical significance levels of the estimated $\hat{\beta}_j$ parameters of equation (1) are unambiguously reflected in the significances of corresponding $APE(x_j)$ indicators. The interpretation of results presented in Table 1 can be outlined using an example based on the age group variable $Age.15.24$: individuals aged 15 to 24 are 4.98% more likely to share their name/surname with private data processors as compared to the reference group of individuals aged 50 and older (we use LP organizers and operators as a common and representative example for commercial data collection & processing). Similarly, our youngest respondent group (15÷24 years) is 23.95% more likely to share email, 15.66% more likely to share address of residence, 26.86% more likely to provide birthdate and 11.43% more likely to share their phone number. All the aforementioned APEs are statistically significant at the 5% significance level – in contrast, there is no significant difference between the propensity to share one's ID number between our examined age group (15÷24) and the reference (50+).

As we move our attention to the age group of 25÷34 years, we can see similar, yet somewhat less pronounced differences in willingness to share personal data (as compared to the reference). Individuals aged 35÷49 years are 4.15% more likely to share name/surname and 18.22% more likely

to share email address compared to the reference age group. Please note the ceteris paribus interpretation: Internet use (and the related possession of any email at all) and other relevant factors are addressed separately in our analysis. Also, propensities to share personal data types (c) to (f) for the group 35÷49 years are not statistically different from the reference. Female consumers (ceteris paribus respondent's age and other relevant factors as in Table 1) are 12.14% more likely to share their address of residence, 8.60% and 8.39% more likely to share their email and birthdate as compared

to men. For the remaining data types, propensities to share are not significantly affected by gender. The inclusion and interpretation of *Educ.Univ* as a regressor to our equation (1) requires separate explanation: although not statistically significant after controlling for all other regressor covariates from equation (1), this variable was consistently selected by the random forest algorithm as an important factor determining personal data sharing attitudes. Therefore, we have included *Educ.Univ* in our main model specification, mainly to ensure consistency in model specification and

Table 1. APEs of selected factors on willingness to share different types of personal data

	NameSurname	Email	Address	Birthdate	Phone	Personal_ID
Age.15.24	0.0498*	0.2995*	0.1566*	0.2686*	0.1143*	0.0299
(HC st.err.)	(0.0211)	(0.0339)	(0.0483)	(0.0481)	(0.0544)	(0.0297)
[p-value]	[0.0183]	[0.0000]	[0.0012]	[0.0000]	[0.0358]	[0.3142]
Age.25.34	0.0423*	0.1955*	0.1194*	0.1337*	0.1211*	−0.013
	(0.0186)	(0.0356)	(0.0474)	(0.0483)	(0.0538)	(0.0242)
	[0.0230]	[0.0000]	[0.0118]	[0.0057]	[0.0245]	[0.5921]
Age.35.49	0.0415*	0.1822*	0.0629	0.0342	0.0216	−0.0266
	(0.0189)	(0.0304)	(0.0436)	(0.045)	(0.0485)	(0.0213)
	[0.0281]	[0.0000]	[0.1496]	[0.4464]	[0.6553]	[0.2112]
Female	0.0341°	0.0860*	0.1214*	0.0839*	0.0523	0.0189
	(0.0179)	(0.0289)	(0.0336)	(0.0342)	(0.0345)	(0.0179)
	[0.0560]	[0.0029]	[0.0003]	[0.0143]	[0.1298]	[0.2926]
Educ.Univ	0.0152	0.0351	−0.0199	−0.0013	−0.0504	−0.0238
	(0.0210)	(0.0342)	(0.0392)	(0.0380)	(0.0377)	(0.0215)
	[0.4692]	[0.3036]	[0.6119]	[0.9718]	[0.1807]	[0.2697]
LS.exotics.yes	0.0526*	0.1071*	0.0613	0.0407	0.0923	−0.0166
	(0.0174)	(0.0363)	(0.0395)	(0.0406)	(0.0399)	(0.0206)
	[0.0025]	[0.0031]	[0.1213]	[0.3170]	[0.0207]	[0.4214]
LS.we.house.no	−0.0362°	−0.0033	0.1745*	0.1115*	0.0621°	0.0533*
	(0.0217)	(0.0313)	(0.0344)	(0.0364)	(0.0366)	(0.0223)
	[0.0962]	[0.9154]	[0.0000]	[0.0022]	[0.0897]	[0.0166]
LS.internet.use.yes	0.0479°	0.1230*	0.0635	0.0457	0.0923*	0.0300
	(0.0255)	(0.0420)	(0.0463)	(0.0451)	(0.0450)	(0.0223)
	[0.0599]	[0.0034]	[0.1705]	[0.3105]	[0.0404]	[0.1788]
LS.cooking.yes	0.0138	−0.0024	−0.0751°	−0.0203	0.0266	0.0197
	(0.0210)	(0.0306)	(0.0391)	(0.0383)	(0.0392)	(0.0233)
	[0.5113]	[0.9383]	[0.0548]	[0.5960]	[0.4970]	[0.3973]
LS.shopping.yes	0.0182	0.0035	0.0561	0.0869°	0.0858°	0.0653*
	(0.0277)	(0.0423)	(0.0461)	(0.0471)	(0.0461)	(0.0308)
	[0.5113]	[0.9337]	[0.2238]	[0.0647]	[0.0631]	[0.0339]
LS.paycard.yes	−0.0436°	0.0297	−0.0409	0.0136	−0.0161	−0.0550*
	(0.0252)	(0.0354)	(0.0412)	(0.0410)	(0.0407)	(0.0228)
	[0.0842]	[0.4023]	[0.321]	[0.7397]	[0.6927]	[0.0159]

Note: * – significant at α = 0.05; ° – significant at α = 0.1.

the validity of our Ignorability of treatment assumption discussed above. Statistically speaking, *Educ.Univ* is marginally relevant for sharing one data type: individual's phone number – but only within the ± 1 standard error interval.

The influences of different lifestyle preferences on personal data sharing propensities are heterogeneous – yet important and statistically significant dependencies can be found. Please note that the term "dependency" is used in statistical context, it cannot be interpreted as a purely causal relationship. Nevertheless, the connection between individual data-sharing preferences and lifestyle factors (hobbies, work and leisure activities) is strong (see also Tahal et al. 2017) and needs to be taken into account (the reasoning is both theoretical: ensuring interpretability through the ignorability of treatment assumption and practical: lifestyle factors are statistically significant in predicting individual data-sharing behavior).

To pick some interesting examples, we provide a brief summary interpretation of the lifestyle factors and their influences on personal data sharing (while the readers are welcome to make their own observations and interpretations based on Table 1): Respondents describing themselves as keen of exotic (seaside) holidays (variable *LS.exotics.yes*) are roughly 10% more likely to share their email and/or phone number as compared to the reference group (less pronounced or dissociative stance towards exotic holidays). Individuals dissociating themselves from owning or using weekend houses (variable *LS.we.house.no*) are 17.45% and 11.15% more likely to share their address of residence and birthdate respectively. At the same time, such consumers are 3.62% less likely to share their name with LP organizers (i.e. less likely to actually share any of their data in order to participate in a LP). While frequent internet users (*LS.internet.use.yes*) are 12.30% more likely to share their email address, they are also 9.23% more likely to disclose their phone number when compared with the reference group. Individuals keen of cooking (variable *LS.cooking.yes*) are 7.51% less likely to divulge their address of residence. For other personal data types, this lifestyle variable does not make a significant contrasting factor. Respondents who report shopping as a significant/important leisure activity are also 8.69% more likely to share their birthdate and 8.58% more likely to provide their phone number. Interestingly, individuals who report being active pay card users are also 4.36% less likely to share their name as compared with the reference group.

Conclusions

In this article, we focus on analyzing personal data sharing preferences from the point of view of a good-faith data processor (LP organizer or other), whose operational environment is changing (perhaps dramatically) and who seeks information relevant for potential/prospective amending of their data-gathering and data-processing activities in a situation where individuals sharing their personal data have an increasing ability and discretion in their decisions to withdraw consent for personal data sharing.

The stratified results of our empirical research provide an actionable toolbox for both academic and professional marketers who seek to identify contrasting factors in individual (consumer) willingness to share personal data. Our research is mainly focused on consumer stances towards actively sharing different types of data with private (commercial, non-government) data processors. Understanding individual motives (statistically speaking) for sharing or not sharing personal data becomes even more important. Many contemporary marketing activities rely on personalized communication, customer segmentation and other data-based approaches. Cost-benefit assessments may lead data processors to fine-tune their data gathering processes (either for LP purposes or similar commercial operations) in favor of personal data providers (consumers), to keep the marketing databases accurate and informative. Alternatively, such cost-benefit analyses may lead to partial redefinition of data gathering and data evaluation processes – based on the amount of personal information that the relevant target group is willing to share.

Finally, our analysis opens multiple suggestions for future research. Once the GDPR-based data protection processes are fully introduced, actual motivations and contrasting factors for requesting information on the extent of personal data collected by data processors may be examined. Also, factors triggering the withdrawal of consent to collect and process individual personal data should be closely investigated, once GDPR-related interactions among natural persons and data processors become habitual.

Author contributions

The authors contributed equally.

Disclosure statement

The authors do not have any conflict of interest.

References

Aldhouse F (1999) The transfer of personal data to third countries under EU directive 95/46/EC. International Review of Law, Computers & Technology 13 (1): 75-79. https://doi.org/10.1080/13600869955260

Ciriani S (2015) The economic impact of the European reform of data protection. Communications & Strategies (97): 41-58 https://ssrn.com/abstract=2674010

Cunningham M (2012) Privacy in the age of the hacker: balancing global privacy and data security law. The George Washington International Law Review 44 (4): 643-695. http://docs.law.gwu.edu/stdg/gwilr/PDFs/44-4/2-%20Cunningham.pdf

EUGDPR (2018) EU GDPR Portal. https://www.eugdpr.org/

Frederik JZB (2015) Personal data processing for behavioural targeting: which legal basis? International Data Privacy Law 5 (3): 163-176. https://doi.org/10.1093/idpl/ipv011

Hastie T, Tibshirani R, Friedman JH (2009) The elements of statistical learning: data mining, inference, and prediction (2nd ed) New York: Springer-Verlag.

International Financial Law Review (IFLR) (1998) Poland: new legislation on the protection of personal data (1998). http://www.iflr.com/Article/1981457/Poland.html

Jankalová M, Jankal R (2017) The assessment of corporate social responsibility: approaches analysis. Entrepreneurship and Sustainability Issues 4 (4): 441-459. https://doi.org/10.9770/jesi.2017.4.4(4)

Kiskis M (2010) Legal regulation of electronic marketing. Jurisprudencija 3 (121): 349-370. https://www.mruni.eu/upload/iblock/a64/20kiskis.pdf

Kosciejew M (2014) Proposing a charter of personal data rights. Information Management 48 (3): 27-31. https://www.questia.com/magazine/1G1-371969300/proposing-a-charter-of-personal-data-rights

Limba T, Agafonov K, Paukštė L, Damkus M, Plėta T (2017) Peculiarities of cyber security management in the process of internet voting implementation. Entrepreneurship and Sustainability Issues 5 (2): 368-402. https://doi.org/10.9770/jesi.2017.5.2(15)

Limba T, Šidlauskas A (2018) Secure personal data administration in the social networks: the case of voluntary sharing of personal data on the Facebook. Entrepreneurship and Sustainability Issues 5 (3): 528-541. https://doi.org/10.9770/jesi.2018.5.3(9)

Majtenyi L (2002) Ensuring data protection in East-Central Europe. Social Research 69 (1): 151-176. http://www.jstor.org/stable/40971541

Manzerolle V, Smeltzer S (2011) Consumer databases and the commercial mediation of identity: a medium theory analysis. Surveillance & Society 8 (3): 323-337. https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/4167/4169

Mitchell A (2010) The rise of volunteered personal information. Journal of Direct, Data and Digital Marketing Practice 12 (2): 154-164. https://doi.org/10.1057/dddmp.2010.28

Murtagh F, Legendre P (2014) Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion?. Journal of Classification 31 (3): 274-294. https://doi.org/10.1007/s00357-014-9161-z

Pounder C (1998) Security and the new Data Protection law. Computers & Security 17 (2): 124-128. https://doi.org/10.1016/S0167-4048(97)81981-5

Reimer K, Becker JU (2015) What customer information should companies use for customer relationship management? practical insights from empirical research. Management Review Quarterly 65 (3): 149-182. http://dx.doi.org/10.1007/s11301-014-0110-z

Robertshaw G, Marr N (2006) Consumer attitudes towards disclosing personal data for direct marketing. Journal of Direct, Data and Digital Marketing Practice 7 (3): 203-215. https://doi.org/10.1057/palgrave.dddmp.4340527

Šišulák S (2017) Userfocus – tool for criminality control of social networks at both the local and international level. Entrepreneurship and Sustainability Issues 5 (2): 297-314. https://doi.org/10.9770/jesi.2017.5.2(10)

Tahal R, Formánek T, Mohelská H (2017) Loyalty programs and personal data sharing preferences in the Czech Republic. E+M Ekonomie a Management 20 (1): 187-199. https://doi.org/10.15240/tul/001/2017-1-013

Tavani HT (1999) Informational privacy, data mining, and the internet. Ethics and Information Technology 1 (2): 137-145. https://doi.org/10.1023/A:1010063528863

Tempest A, Brandau G (2010) Data protection in Europe – a cloudy future. Journal of Direct, Data and Digital Marketing Practice 12 (2): 180-193. https://doi.org/10.1057/dddmp.2010.30

Wooldridge JM (2010) Econometric Analysis of Cross Section and Panel Data (2nd ed) Cambridge, Massachusetts: MIT Press.

Zeman T, Břeň J, Urban R (2017) Role of internet in lone wolf terrorism. Journal of Security and Sustainability Issues 7 (2): 185-192. https://doi.org/10.9770/jssi.2017.7.2(1)

APPENDIX

It provides explanation of the process involved in using Likert scale-based variables describing lifestyle factors for subsequent quantitative analysis. In the questionnaires, we use questions with a 5-degree Likert scale answers in order to assess individual lifestyle preferences. Likert scale variables are ordered categorial data, bearing information on the ordering only (say, '1' is better than '3'). Although basic descriptive statistics (such as mean values) are commonly provided in literature, we cannot assume underlying linear nature of the surveyed Likert scale data, necessary for direct use in regression analysis (we cannot say that the difference between '2' and '4' is twice the magnitude of the difference between '1' and '2'). Therefore, our methodology of recording the answers for further analysis is based on binary variables. The process may be described using a simple example: In the questionnaire, respondents are presented with a statement "I use the Internet frequently." For this statement, respondents are asked to position themselves on a five degree Likert scale ('1' = this statement describes me very well, … , '5' = this statement does not describe me at all). Subjectively perceived attitude is addressed here, rather than actual time spent, data volumes, etc. Answers to this question are used to generate two variables: $LS.internet.use.yes_i$ equals 1 for respondents who report '1' on the Likert scale and zero otherwise, $LS.internet.use.no_i$ = 1 for individuals dissociating themselves from the statement by answering '5' (and it equals zero otherwise). This allows for identification of individuals with a clear and strong stance towards a given subject. All the remaining Likert scale answers ('2' to '4', i.e. not a very strong position on the subject) are implicitly combined into one reference category that is used for proper result interpretation.

Notations

Variables and functions

$Age.15.24_i$ – binary variable, equals 1 if the i-th respondent is aged 15÷24 and is zero otherwise;

$Age.25.34_i$ – binary variable, equals 1 if the i-th respondent is aged 25÷34 and is zero otherwise;

$Age.35.49_i$ – binary variable, equals 1 if the i-th respondent is aged 35÷49 and is zero otherwise (two additional age groups: 50÷64 and 65+ exhibit homogenous behavior and thus have been combined into a single reference group);

$Female_i$ – distinguishes respondent's gender, equals 1 for women;

$Educ.Univ_i$ – discerns individuals with a university degree (basic and secondary education form a single reference category);

$LS.internet.use.yes_i$ – Likert scale-based lifestyle variable. Indicates whether or not individuals identify themselves strongly with a frequent use of the Internet (subjectively perceived attitude is addressed here, rather than actual time spent, data volumes used, etc.). Technical details on Likert scale-based variables are provided in the Appendix;

$LS.exotics.yes_i$ – Likert scale lifestyle variable, indicates whether or not individuals decidedly identify themselves as keen of exotic and sea-side holidays (the Czech Republic is a landlocked country) – both active participation and aspirational attitudes are addressed here;

$LS.w.e.house.no_i$ – Likert scale-based, marks respondents who strongly dissociate themselves from using/owning a weekend-house (cottage);

$LS.cooking.yes_i$ – equals 1 for individuals fond of cooking and zero otherwise;

$LS.paycard.yes_i$ – identifies frequent pay card users (individual attitudes are addressed, rather than money spent);

$LS.shopping.yes_i$ – distinguishes individuals who enjoy shopping as a leisure activity (individual attitude addressed);

R_i – regional effect (similar to the effect as in Figure 1) that we control for in order to obtain consistent partial effects of the sociodemographic and lifestyle factors described above. Its construction, estimation and testing follow the panel-data 'unobservable' effect methodology for fixed effects analysis (Wooldridge 2010);

u_i – random element of the regression model;

y_i – dependent variable of the regression model;

$\beta_j, \hat{\beta}_j$ – j-th regression parameter of the model: population and sample-based estimate;

n – sample size.

Abbreviations

APE – Average partial effect;

CRM – Customer relationship management;

GDPR – General Data Protection Regulation – regulation (EU) 2016/679 of the European Parliament;

HC st. err. – Heteroskedasticity-consistent standard error;

LP – Loyalty program;

NUTS2 – level 2 of the EU's classification of territorial units for statistics (abbreviation comes from French: Nomenclature des Unités Territoriales Statistiques).