Reflection of GDPR by the Czech Population

Radek TAHAL

*University of Economics, Prague, Czech Republic
radek.tahal@vse.cz*

Tomáš FORMÁNEK

University of Economics, Prague, Czech Republic

Abstract. *Nowadays, personal data is becoming both increasingly relevant and extensively shared over the internet. Individual privacy is often at risk as personal and contact information are a frequent object of commercial transactions. This situation recently triggered the implementation of a new data-protection regulation: General Data Protection Regulation (GDPR). This study analyses privacy protection from the citizens' and consumers' points of view. The authors apply quantitative methods to primary data sampled from the population of the Czech Republic (aged 15+). Individual willingness to share personal data is evaluated along with the spectrum of situations where such data is requested and along with corresponding GDPR-related individual stances. According to the data analyzed, Czech population perceives GDPR measures as mostly useful yet somewhat bothersome. Also, individuals are not convinced that the regulation could deliver any fundamental changes to processes that involve handling of personal data. In general, respondents feel concerned by the intensification of demands for signing data sharing consents and by similar data processing interactions. However, individual levels of willingness to share various types of personal and contact information are sound and stable.*

Keywords: personal data, GDPR, quantitative analysis, Czech population, data protection.

Please cite the article as follows: Tahal, R. and Formánek, T. (2020), "Reflection of GDPR by the Czech Population", *Management & Marketing. Challenges for the Knowledge Society*, Vol. 15, No. 1, pp. 78-94. DOI: 10.2478/mmcks-2020-0005.

Introduction

Abundant new on-line communication technologies and services bring many benefits to their users, often without a direct price tag to the consumer. However, this is often at the cost of individual privacy. Huge datasets are created and managed and they often contain personal information. Such data is generated on social networks, as consumers purchase goods or services on-line, by contacting a local authority office, by visiting a doctor or in connection with employment or study undertakings. Collected personal data are often very sensitive and may become a target of data theft or other potential misuse.

It is the increasingly frequent storage, use and sharing of personal data that launched a wide-range discussion on the steps necessary to protect EU's citizens from the possibility of uncontrolled sharing, publication and misuse of personal data. The outcome of this discussion was the introduction of a new directive called General Data Protection Regulative (GDPR, regulation EU 2016/679), which unifies conditions and sets rules for the use of personal data throughout the EU. This directive came into force on May 25, 2018 and it set rules for personal information handling that apply to both private and public entities. It should be noted that GDPR legislation distinguishes (Article 4, "Definitions") between

data controllers, data processors and recipients: data controllers are different types of entities that determine the purposes and means of the processing of personal data and data processors are any entities (natural or legal person, public authority, agency, etc.) that work with personal data on behalf of the controllers. Recipients are entities (controllers or third parties) to which personal data is disclosed (as a general rule, appropriate public authorities are not regarded as recipients).

The main protective purpose of the GDPR regulation is accomplished by restricting data processors to work with personal and contact data only based on explicit and revocable individual consent. Importantly, data processors must clearly declare for what purpose the information is used, how data is stored and secured.

This study evaluates and describes personal data protection topics as perceived by individuals: citizens and consumers, whose protection is enforced by the new GDPR legislation.

Literature review

Data protection is a global issue and a topical concern to many data-handling entities. The potential for misuse of sensitive personal information originates from both state and commercial entities that have the capabilities to accumulate and use personal data improperly, perhaps in violation of personal rights (Abdulrauf and Fombad, 2017).

Data protection possibilities and potential bottlenecks have been discussed for a long time. This topic even predates the massive onset of the internet. The discussion, however, was greatly accelerated by digitalization of social interactions and by globalization of the economy. Nyoni and Velempini (2018) discuss different aspects and threats to individual privacy on social media and other customer/consumer databases that are subject to regular commercial transactions involving structured and focused advertising activities, etc.

To describe why the new GDPR was implemented, Moerel (2011) characterizes the preceding legislative framework by pointing out that Data Protection Directive (DPD) predates the contemporary wide use of the Internet. Hence, DPD and the regime it established were not devised with the vast increase in cross-border flows of personal data in mind – a fact that resulted in the introduction of the new regulation. As Helberger et al. (2017) point out, the increasingly frequent commoditization and monetization of personal data has contributed greatly to the discussion on the protection of personal data. In the on-line world, many companies offer “free” services in order to monetize consumer data which they collect through such services (see e.g. Ruggieri et al., 2018).

This topic is also reflected by Carugati (2018, p. 4): “To obtain data, one can develop a data processor or simply buy data from a third party, such as a data broker. Another path is to simply acquire a company with colossal amounts of data.” While the “simplicity” of acquiring companies may be questionable, the commoditization and asset-like nature of personal data is indisputable. Hribar et al. (2018) point out two prominent and rapid effects of GDPR: Personal data protection is dramatically strengthened and easier to control; at the same time, data controllers and processors are subject to more stringent duties and legal obligations.

Kramer (2002) compares personal data protection in two main areas (markets): while the EU authorities chose to protect personal information through legislation (i.e., GDPR), the United States decided to protect consumer privacy through market mechanisms

and the promotion of industry self-regulation. Koops (2014) underlines that every citizen should be able to control the personal data they provide to someone. His “Informational self-determination” is derived from the fact that people should be able to exercise control over what happens with their personal data. According to this concept, individuals’ free and informed consent is an important ground to legitimize data processing. “Information exchange is based on an expectation of the transparent sharing of timely and useful information” (Voges, 2014, p.387). Also, it implies various corresponding controls over the data, such as rights to correction or erasure (see also Articles 16 and 17 of the GDPR).

Different personal data protection aspects are closely connected to the Big Data phenomenon: for many data processors (social networks, on-line retailers, etc.), sizes and complexities of data sets may go far beyond the ability of commonly used software to record, process and use (interpret) data within a tolerable amount of time. Here, data processing often involves drawing data from a potentially wide variety of datasets that, historically, were never intended to be combined. (Murphy and Barton, 2014). Veghes et al. (2012) elaborate on contemporary topics of privacy and personal data protection, associated with consumer data being collected on-line. Similarly, Crabtree et al. (2016) analyze the Big Data and Internet of Things (IoT) phenomena, as they constitute a new paradigm to an economic market that is increasingly driven by personal data.

As soon as the GDPR directive came into force, its practical effects (both positive and negative) may be observed. As noted e.g. by Cvik et al. (2018), data processors were facing multiple issues concerning accurate and smooth implementation of this new regulation. This is only logical, as the regulation covers many new topics, most of those being highly technical (<http://www.privacy-regulation.eu/en/index.htm>).

Personal privacy protection must be addressed in context with the particular area where it is applied. (Veghes et al., 2009). The fact that personal data protection topics are a concern to be addressed by both business entities and governmental institutions (participating in public affairs and state administration) is addressed e.g. by Korze and Certanec (2017) who point out that such data is generated and processed by many different public entities, ranging from law enforcement, other rapid response operators, healthcare providers, etc.

It is desirable that, on the one hand, the protection of privacy serves its purpose, but on the other hand, it should not impose excessive burdens on peoples’ lives. “Consumer’s privacy rights may not be overridden, since their positive acceptance of and active use and involvement in the new technology are key concerns and parameters for achieving its very creditable goals.” (Knyrim and Trieb, 2011, p.121).

In one of the most recent GDPR reflections in scientific literature, Hoofnagle and Borgesius (2019) point at the manifold implications involved: while the regulation complicates and restrains some information-intensive business models, it also enables and promotes alternative approaches that might have been deemed ineffective under the previous less-protective environment. GDPR may even promote world-wide technological development. To minimize liability under the new EU’s regulation, organizations around the world need to make changes to comply with the GDPR (Li et al., 2019).

Methodology

The above-discussed data protection issues are at the center of this analysis. A relatively broad range of GDPR-related topics may be formalized into three research questions (RQ):

RQ1: Do individuals think GDPR is a useful regulation and in what situations do they regard it as beneficial?

RQ2: In what situations have individuals encountered a requirement to sign personal data sharing consent? Do they consider this request annoying?

RQ3: To what extent are individual stances on GDPR regulations influenced by socio-demographic factors? How pronounced are the observed gender, education and age-based differences?

To address the above RQs, this article is based on results from multiple quantitative analysis approaches applied to primary data. Data is sampled from the population of the Czech Republic (age 15+). Sampling was performed using quota selection, taking into account three factors: location, age and gender of respondents. Data collection was performed from October to December 2018. The survey was organized and carried out by a research team at the University of Economics, Prague. The research team was 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 scientific surveys and analyses focused on personal data protection and GDPR (see Tahal et al., 2017 or Formánek and Tahal, 2018).

Complete and valid response-sets were gathered from a total of 477 respondents. The survey data collection process combined both personal interviews (subsequently stored on-line by interviewers) and directly accessed on-line questionnaire (the Czech language-based questionnaire is available for reference at https://docs.google.com/forms/d/1aXnS7kZVpbTG0BtDpkmWvAusTnnPmMylEfP9O5v4_Co/edit?chromeless=1). The questionnaire consists of 15 “main” closed-form questions (some of the questions encompass sub-topics). The survey was focused on gathering respondents’ sociodemographic data (4 questions) and answers to questions related to personal data protection and GDPR stances. Different types of questions were used in the survey: quantitative (mostly interval-based), qualitative (Yes/No) and Likert scale (5-degree span is used). In total, the questionnaire-based data was transformed into a dataset with 477 rows/observations and 52 columns/variables for subsequent quantitative analysis. Survey data-validation was performed using the Wald-Wolfowitz “Runs” test (see Wackerly et al., 2008), where the null hypothesis of random ordering of observations is tested against the alternative of potential survey mishandling.

Questionnaire construction and the variation of questions are in accord with two preceding qualitative research operations focused on a similar topic: a group discussion held on 23rd October 2017 at KANTAR (a research and consulting private entity) and personal in-depth interviews therein with individuals from the general population of the Czech Republic (October - November 2017). Both those researches were conducted in co-operation with a research team at the University of Economics, Prague.

In this paper, we analyse two interrelated aspects of the GDPR regulation: its effectiveness and its popularity in various circumstances. Also, we examine the extent to which socio-demographic differences affect individual stances on GDPR. As our observed variables are generated from Likert scale-based questionnaires, some limits to quantitative analysis need to be observed.

For example, let’s consider a 5-element Likert scale (value “1” = worst outcome; ... ; value “5” = best outcome). The fact that “5” is a better rating than “4” conveys useful information, but the rating has ordinal meaning only - we cannot say that the difference

between “3” and “5” is twice as important as the difference between “4” and “5”. On the other hand, mean values are commonly used to describe the “central tendency” of Likert scale measurements, as opposed to the formally more appropriate frequency and/or contingency tables. Saal et al. (1980) argue in favor of Likert scale mean values: those can be easily calculated, they are simple (single number) and intuitively understood by the general public. Last but not least, if Likert scale-based means are compared (among populations), they usually yield the same rank orderings as other indices (e.g. median values).

However, this “mean simplification” may not be easily extended to other statistics, such as covariance/correlation coefficients. As we observe Likert scale data (ordered multinomial variables), polychoric correlation coefficients should be used. Theoretical overview of the subject is provided next.

Polychoric correlation

The principle of polychoric correlation dates to the work of Pearson (1900) who pointed out that some variables may be very difficult to measure on a quantitative scale, yet their individual observations may be easily placed on a scale of ordered categories. Polychoric correlation is a technique for estimating the correlation between two ordered multinomial variables. This method usually assumes the existence of underlying (unobserved) normally distributed continuous variables that generate our ordinal observations.

Say, we have two ordered multinomial variables A and B that are related to their underlying (latent) counterparts X and Y so that

$$\begin{aligned} A = a_i & \quad \text{if} \quad \gamma_{i-1} \leq X < \gamma_i, \quad i = 1, 2, \dots, r \quad \text{and} \\ B = b_j & \quad \text{if} \quad \tau_{j-1} \leq Y < \tau_j, \quad j = 1, 2, \dots, s, \end{aligned}$$

where γ_i are thresholds given at the scale of latent variable X and a_i is the i th category of A (notation is analogous for B and Y). Both the thresholds and values of corresponding categorical variables follow a strictly increasing order with $\gamma_1 < \gamma_2, \dots, < \gamma_{r-1}$ and $a_1 < a_2 < \dots < a_r$ (analogous for τ_j and b_j) and we assume $\gamma_0 = \tau_0 = -\infty$ and $\gamma_r = \tau_s = \infty$. Without a loss of generality, we may set the mean and variance of for X and Y as $\mu = 0$ and $\sigma^2 = 1$. Using a product-moment correlation between X and Y of $\rho = \text{cov}(X, Y)/\sigma$, we may cast the standard normal bivariate density of X and Y as

$$\phi(x, y; \rho) = \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp \left[\frac{-1}{2(1-\rho^2)} (x^2 - 2\rho xy + y^2) \right], \quad (1)$$

where x and y are observations of variables X and Y . Now, the probability of simultaneously observing $A = a_i$ and $B = b_j$ can be formalized as

$$P_{ij} = \int_{\gamma_{i-1}}^{\gamma_i} \int_{\tau_{j-1}}^{\tau_j} \phi(x, y; \rho) dy dx. \quad (2)$$

If we denote n_{ij} the number of observed realizations such that conditions $A = a_i$ and $B = b_j$ hold simultaneously in a sample of $n = \sum_i^r \sum_j^s n_{ij}$ (pairs of observations), we can formalize the likelihood function of the sample as

$$L = K \prod_{i=1}^r \prod_{j=1}^s P_{ij}^{n_{ij}}, \quad (3)$$

where K is a constant. Maximum likelihood (ML) estimation of the polychoric correlation (sometimes denoted as ρ_{pc}) is produced by log-transforming the likelihood function (3), differentiating with respect to all parameters $(\rho, \gamma_1, \dots, \gamma_r, \tau_1, \dots, \tau_j)$, equating all partial derivatives to zero and solving the equation system. Alternative (computationally less expensive) estimation algorithm is provided by Martinson and Hamdan (1971) and the R implementation by Fox (2016) can handle deviations from normality assumptions for the bivariate distribution of X and Y .

Polychoric correlations are on the scale of $\langle -1, 1 \rangle$, with the usual interpretation. Also, we need to bear in mind that correlation is not causation. Hence, ρ_{pc} cannot be interpreted as if one variable “causes” the other. A closely related tetrachoric correlation is a special case of ρ_{pc} , with A and B variables being observed on a binary $\{0, 1\}$ scale. Similarly, polyserial correlation would relate to the case where one of the variables is observed on a continuous scale while the other is polychotomised (Drasgow, 1986).

Given the discrete and categorical nature of our observed variables, polychoric correlations provide an effective way to assess homogeneity/non-homogeneity of GDPR-related individual stances. Compared to the usual χ^2 tests of homogeneity (see e.g. Wackerly et al., 2008) – where we would only reject/not reject the null hypothesis of homogeneity (alternative null hypotheses are possible yet rarely practical) – polychoric correlations provide additional interpretation possibilities. Besides homogeneity conclusions (if polychoric correlation is not statistically different from zero when measured e.g. between a gender based dummy variable and a Likert scale-based stance on some GDPR aspect), polychoric correlation coefficients provide information on the sign and strength of the relation between such variables. Also, while the χ^2 tests of homogeneity can only be used for different populations (male/female, age groups, etc.), polychoric correlation may be computed much more generally – to compare correlation between categorical answers to survey questions, etc.

Research results

To answer our RQ1 to RQ3, we start by focusing on GDPR benefits given by its personal data protecting nature. Respondents provided their stances on GDPR usefulness in four selected aspects related to sharing personal data during common interactions with state authorities, local authorities (municipalities) and businesses. A five-element Likert scale was used to express respondent’s agreement with GDPR usefulness: 1 is the least favourable position: “GDPR does not help at all” and 5 indicates that “GDPR is actually very helpful”. A brief summary of the results is shown in Table 1.

Table 1 reflects a relatively high degree of respondents’ indecisiveness in terms of their opinion on the benefits of GDPR regulation. Overall, individuals are somewhat inclined to view GDPR favourably, because it yields additional control and confidence into the

information provided to data processors – see topic/question (d). As perceived by respondents, GDPR helps in preventing potential misuse of personal information related to data leakages, thefts, etc. – topic/question (b).

Table 1. GDPR usefulness in specific tasks involved with processing personal data

GDPR-induced effect	Average evaluation *
(a) Greater personal data security for people	2.82
(b) Preventing leakage of sensitive information	3.23
(c) Data processors follow rules when dealing with data	2.68
(d) Individuals can control what information is shared	3.21

*1: worst evaluation, 5: best evaluation

Source: Authors' own research.

Following upon the information presented in Table 1, the perceived usefulness of GDPR in the above discussed situations (a) to (d) was further investigated with respect to any potential socio-demographic differences. In Table 2, the responses to topics/questions (a) to (d) are stratified along distinct gender, age and education groups. As situations (a) to (d) are aggregated in Table 2, the summary ("averages") are obtained from the same five-element Likert scale described above.

Overall, the perceived usefulness of GDPR may be represented by an average value of 3.08 on the five-element Likert scale. This corresponds to the middle value of the scale used. It is worth mentioning that each of the extreme values, i.e. "1" and "5" were selected by 9 % of respondents.

Table 2. Socio-demographic aspects of the perceived GDPR usefulness

Population Group	Mean evaluation *
Population Total	3.08
Male	2.91
Female	3.17
Age 15 – 26	3.15
Age 27 – 35	3.01
Age 36 – 48	3.14
Age 49 – 65	2.68
Age 65 +	2.60
Primary education	3.31
Secondary school without final state exam	3.27
Secondary school with final state exam	3.07
University degree	2.97

*1: worst evaluation, 5: best evaluation

Source: Authors' own research.

The analysis provided in Table 2 shows that positive stances on the usefulness of GDPR regulation are slightly decreasing with respondents' age and with their higher education. It can be stated that women and younger individuals (aged at or below 48) consider GDPR as more beneficial. The same result holds for respondents who acquired lower education levels. In contrast, men, older respondents, and people with higher education perceive GDPR less favourably.

Tables 3 to 5 provide additional stratified insight into dependencies between respondents' positions on GDPR necessity (justification) and its actual usefulness/efficiency. Table 3 describes polychoric correlation (ρ_{pc}) for individual positions on GDPR efficiency, its actual necessity (justification) and respondent's age (respondent's gender was found statistically insignificant – i.e. it is a non-differentiating factor for this type of individual stances). Specifically, using a 5-element Likert scale (1: “GDPR does not help at all, ... , 5: “GDPR is actually very helpful”), individuals were asked to provide their position on the efficiency of GDPR protection in the following areas: Q10-A: Increased individual safety, Q10B: Stopping sensitive information leakages/thefts, Q10C: Legal compliance and Q10D: Control over individual's personal information sharing. In a separate question (Q11), respondent were asked to evaluate GDPR's necessity (1: “Totally unnecessary”, ... , 5: “Very necessary”). Responses to questions Q10A-Q10D and Q11 were compared to respondents' age (5-level ordered scale as shown in Table 2).

Table 3. Polychoric correlations: GDPR efficiency and necessity

Source: Authors' own research.

From Table 3 (symmetric along main diagonal by construction), we can see that individual stances on GDPR usefulness/efficiency in different areas of action are positively correlated, with a set of relatively uniformly distributed ρ_{pc} coefficients, ranging from 0.45 to 0.50. Here, correlation between individual safety (Q10A) and sensitive information (Q10B) protection is the highest, with $\rho_{pc} = 0.67$. Correlations between GDPR necessity

(justification) and efficiency in different fields of action are observed within a similar range (0.43 to 0.48) with an interesting exception of $\rho_{pc} = 0.37$ for personal information sharing control (Q10D) and GDPR necessity (Q11). As we focus on the last row (column) of Table 3, we can see that older individuals (as compared to younger respondents – see also Table 2) tend to consider GDPR to be less effective. This negative relationship is strongest for individual stances on sensitive information leakage (Q10B) with $\rho_{pc} = -0.27$, while there is no statistically significant relationship between age and perceived GDPR efficiency with respect to control over personal information sharing (Q10D).

As discussed in the methodology section, polychoric correlation and corresponding statistical inference assume normality of the underlying (latent) variables. As this assumption may not hold for our dataset (if common normality tests such as Jarque-Berra are applied to the latent variable “observations” extracted from the polychoric correlation calculation algorithm), we provide bootstrapped (non-parametric, i.e. empirically derived) 95% confidence intervals (CIs), calculated based on the R routines available from Revelle (2018).

Table 4. Polychoric correlations and bootstrapped 95% confidence intervals

	lower.emp	lower.norm	estimate	upper.norm	upper.emp	p-value
Q10A-Q10B	0.59	0.59	0.67	0.72	0.72	0.00
Q10A-Q10C	0.40	0.39	0.50	0.59	0.58	0.00
Q10A-Q10D	0.36	0.34	0.44	0.53	0.53	0.00
Q10A-Q11	0.40	0.39	0.48	0.57	0.56	0.00
Q10A-Q13	-0.30	-0.32	-0.22	-0.11	-0.11	0.00
Q10B-Q10C	0.31	0.33	0.44	0.53	0.54	0.00
Q10B-Q10D	0.41	0.39	0.49	0.57	0.56	0.00
Q10B-Q11	0.32	0.31	0.43	0.53	0.51	0.00
Q10B-Q13	-0.34	-0.35	-0.27	-0.16	-0.17	0.00
Q10C-Q10D	0.36	0.37	0.45	0.53	0.52	0.00
Q10C-Q11	0.37	0.38	0.46	0.54	0.53	0.00
Q10C-Q13	-0.22	-0.23	-0.14	-0.06	-0.05	0.00
Q10D-Q11	0.29	0.29	0.37	0.46	0.45	0.00
Q10D-Q13	-0.15	-0.15	-0.07	0.02	0.04	0.13
Q11-Q13	-0.20	-0.21	-0.12	-0.01	-0.01	0.03

Source: Authors' own research.

Table 4 expands on the information provided in Table 3 by comparing normality-based CIs to their bootstrapped counterparts that are free from distributional assumptions. Based on Table 4, we may see that the differences between normality-based CIs and bootstrapped CIs are almost negligible (no effect on overall statistical significance at the $\alpha = 0.05$ level (for additional discussion on this topic, see Lumley et al., 2002). Please note that for identification and readability, Table 4 features the ρ_{pc} coefficients – repeated from Table 3. The last column of Table 4 is used for testing statistical significance: the null

hypothesis of statistically non-significant correlation is rejected at $\alpha = 0.05$ for all correlations, except for the Q10D-Q13 row. By comparing Tables 3 and 4, one can see how statistically insignificant results are printed in Table 3 using a very small font size (sign distinction is made by color-coding).

From the output as presented in Table 5, we may observe correlations among the general fields where our respondents reported to interact with GDPR-related consent requests. Q5 is a Yes/No (1/0) question focused on whether an individual was asked to provide consent with sharing personal data “lately” (to avoid respondents’ distraction, no specific time-test was implemented here). The following questions are more specific – focused on different types of social environments: Q6A relates to GDPR-related consent being requested from the respondent at work, Q6B is at school, Q6C in a medical facility, Q6D in a shop (consumer-market related environment), Q6E for the government authorities (local/central) and Q6F relates to GDPR consent requested by a bank. Q7 is another binary-scale question (Yes/No), aimed at finding out whether respondents perceive GDPR consent requests as annoying. Q13 follows from Table 3 and reflects population’s basic age structure (we found no statistically significant gender-based correlations and the corresponding row/column was excluded from Table 5).

Table 5. Polychoric correlations – GDPR consent requests and interactions

Source: Authors’ own research.

The results in Table 5 show a few interesting and interpretable results. Economically active individuals (only such individuals may become subject to a GDPR-consent request at work – i.e. under Q6A) tend to interact with GDPR-related issues in other non-work areas – presumably this is given by the fact that they lead more dynamic and vivacious lives: positive correlation for questions Q6D, Q6E and Q6F coincides with economically active individuals being more active overall: they also shop more, interact with authorities for

various reasons and purposes as well as use banking services more. The highest correlation was observed for GDPR consent requested by government authorities and banks: $\rho_{pc} = 0.62$ for Q6E and Q6F. From the last row/column of the table, we can see that older people tend to interact less with different types of GDPR-related issues. Besides the overall negative correlation between age and GDPR-consent requests, we may use Table 5 for a stratified analysis. The negative correlation between age and GDPR requests at school is logical. Please note the positive correlation for age and shopping (Q13 and Q6D) – presumably, older people tend to participate more in loyalty programs where personal information (and thus GDPR consent) is usually required.

Interestingly, we find no correlation between medical related GDPR-consent requests and other types of requests (Q6C vs. other Q6 questions). Please note this is not due to e.g. small prevalence of GDPR-related interactions in some of the categories as the prevalence of reported GDPR-consent requests ranges from 33% (Q6E) to 52% (Q6A). The same conclusion would apply to GDPR at school (Q6B) but this outcome is slightly correlated with shopping (Q6D) at $\rho_{pc} = 0.15$. Also, using the same approach as in Table 3, statistical significance of all results in Table 5 can be evaluated by the font size used (normal size for significant results).

In the questionnaires, respondents (representative members of the general public) were asked to identify environments and/or situations in which they were asked to provide personal data sharing consent and whether they considered this data consent procedure bothersome. Interestingly, only 3 % of individuals reported no previous experience with GDPR-related consent requests. The remaining 97 % of respondents stated that they had encountered such request at least once in the environments/situations listed in Table 6. The list of environments included in the questionnaire follows directly from the analysis of previous qualitative research activities, as described in Section 2.

Table 6. Personal data consent request – environment/situation overview

Environment of consent request/interaction	Percent Respondents
a) At work	52 %
b) At school	39 %
c) In a medical facility	40 %
d) In a retail shop	34 %
e) Public administration in general	33 %
f) In a bank	42 %
h) Other	12 %

Source: Authors' own research.

Table 6 summarizes the reported relative frequencies of environments/situations in which members of the general public interacted with GDPR-related requests. Following the nature of the phenomenon observed, respondents were presented with a multiple-choice question, to identify different environments as appropriate.

Category "Other" (h) was supplemented by an open question asking for specific description of the environment/situation involved. Here, respondents' most frequent

answer pointed at personal information consent being given “on the internet”, often specifying internet-based purchases of goods or services.

Figure 1 expands on the information presented in Table 6 by providing structured insight into respondents’ sentiments towards the GDPR-related consent process. This part of analysis specifically addresses whether individuals perceive (describe) the process of data sharing consent provision as bothersome.

The findings reflected in Figure 1 may be briefly summarized as follows: middle-aged individuals (36 - 48 years) and respondents with a lower education level (secondary education without final state exam) are significantly more likely to perceive signing/processing the GDPR-based consent as inconvenient (bothersome). Interestingly, this sentiment is not repeated in the “Primary education” group. This is mostly caused by the fact that “Primary education” group is predominantly saturated by individuals aged 15 to 26; many of them have not reached higher education yet, simply because of their age.

Figure 1. GDPR-based consent process perceived as bothersome

Source: Authors' own research.

Next, we focus on evaluating individual levels of willingness to share different types of personal and contact information. This generalizes and expands our previous analyses on consumer attitudes towards providing personal and contact information to retailers for the purpose of loyalty program participation (Formánek and Tahal, 2018; Tahal et al., 2017).

For eight different types of personal and contact data, respondents were asked to describe their level of willingness to share information. Such willingness was described using a five-element Likert scale, with “1” representing the lowest level of willingness to share data and “5” representing the highest level. Figure 2 summarizes the results of this part of our analysis.

From Figure 2, we can see that respondents are able to differentiate well between personal data types and their sensitiveness levels. Name is a basic individual identifier, with little misuse potential (unless combined with other types of data). This is reflected the highest overall willingness to share this type of data, observed at an average level of 3.9 on

the five-element scale described above. E-mail (and nationality information) follow closely at 3.4 and the date of birth (i.e. age information) scored at 2.7 points. Please note how protective individuals are towards their phone numbers and address of domicile information (2.5 and 2.3 point averages respectively). Such observed results – personal data sharing attitudes – confirm some of our previous findings (e.g. Tahal et al., 2017). Namely, respondents consider their phone number as a confidential information and they are not willing to share this type of personal/contact information, much in opposite to their e-mail address. Personal ID is the most protected data type. Actually, it serves more as a benchmark for evaluating propensity to share other types of data. In the Czech Republic, most data processors are not allowed to legally request and store (use) this type of data under most practical circumstances – exceptions such as insurance services and airline tickets may apply.

*1: lowest willingness ; 5: highest willingness to share personal data

Figure 2. Willingness to share different types of personal data

Source: Authors' own research.

Limitations

A complete enumeration of situations/environments where individuals are asked to provide personal data is virtually impossible. For example, many different sub-variants or generalizations exist for the option (d) "In a retail shop" as in Table 6. Customer's GDPR-related experiences may be significantly different based on retail store size and focus. Registering for a loyalty program managed by a retail/supermarket chain would be in a significant contrast when compared to registering with a regional or family-run business. While we cannot cover all possibilities and nuances of personal data sharing interactions, our results may efficiently outline the general picture: we have mapped current stances of the Czech population on principal GDPR topics, based on a survey that was performed six months after the legislation came into force.

When addressing sensitivity of e-mail addresses (and individual willingness to share them), we need to bear in mind that many individuals possess several such electronic contacts (e-mails) and they often use them for various purposes. Also, many "free-mail" addresses (offered as free service by multiple on-line entities) are viewed as preserving user anonymity to a great extent. Those may be the main reasons why providing e-mail address is significantly less sensitive when compared to domicile address or phone number. Such conclusions are based on the group discussions described in Section 2 as well as

supported by previous researches. For example, Woo (2006) points out that secrecy and deception regarding identity might be the most effective way of ensuring on-line privacy. Following the same logic, unsolicited phone calls rank among the most annoying activities in terms of intrusion into individual privacy. This finding is implied by both the questionnaire-data and by indepth interviews performed and supported by previous researches (e.g. Formánek, Tahal, 2018).

Many possibilities exist for data processors to request personal data sharing consent. From a simple checkbox within an electronic application, over-the-phone consent given during a telemarketing interaction or even the traditional filling and signing of a paper form. Hence, individual stances on the level of inconvenience involved with providing GDPR-related consent as presented in this contribution are necessarily somewhat generalizing.

The data survey for our study was carried out roughly 5 to 6 months after the GDPR regulation came into force on May 25, 2018. While the new legislation was widely discussed in Czech media and most of the respondents reported first-hand experience with GDPR and/or personal data protection concerns, it would be beneficial to repeat the survey after a moderate time period. Additional insight might be gained by surveying individual stances once the general public and data processors have enough time to adjust their behavior and once GDPR's real-time strengths and weaknesses become apparent.

Discussion and conclusions

The new EU's GDPR legislation entered into force on 25 May 2018. This regulation has been subject to large-scale debates ever since it has been passed, as it imposes the most comprehensive set of data protection rules world-wide. GDPR affects virtually all subjects collecting and processing personal data. Businesses (especially small and medium-sized companies) and public institutions have sometimes struggled significantly to comply with the regulation in all its aspects. Many data processors had to ask their customers (students, patients, etc.) to sign new/amended versions of personal data sharing consents. Many consumers (and individuals in general) were often unpleasantly surprised by receiving calls or written requests for signing new, updated, business (other) conditions to comply with GDPR. This contribution focuses on attitudes of the Czech population towards the new GDPR legislation, as expressed approximately six months after its full implementation. Specifically, primary data for our analysis were collected during the period October to December 2018.

Our analysis shows that the Czech population perceives GDPR regulation with certain amount of reservation, given the bothersome aspects involved with the procedure of providing data sharing consent. On the other hand, benefits are visible as GDPR makes individuals think more about the types of information they want to share and about the different situations/environments where personal data is requested. Also, GDPR is viewed as providing more confidence into diligent and responsible data management being observed by data processors. However, GDPR is not viewed as providing any fundamental changes to basic data security.

Our research reveals a sound stability of attitudes towards the willingness of sharing various types of personal and contact data. The authors of this contribution had carried out and published personal-data related analyses even before the GDPR was implemented. This contribution confirms and enhances the scope of conclusions obtained

through two preceding studies (Tahal et al., 2017, Formánek and Tahal, 2018). GDPR legislation does not change individual willingness to share different types of personal data. There are hardly any adverse sentiments involved with sharing one's name or e-mail. On the other hand, phone number and address of domicile are among the most sensitive data types.

While this article provides a complex and structured analysis of GDPR-related stances expressed by the Czech population, various topics remain open for future research. As communication technologies and on-line services are evolving rapidly, so do the possibilities of personal data use/misuse. Social networks and other data processors may have amended their procedures to conform with GDPR, yet the economic incentives involved with personal data handling and mining remain unaffected. Hence, this type of research (surveys and in-depth interviews) should be repeated periodically to evaluate perceived GDPR effectiveness and usefulness. Ideally, this type of surveys might be performed in various EU countries, to assess common views and gain additional insight by investigating possible differences.

References

- Abdulrauf, L. A., & Fombad, C. M. (2017). Personal data protection in Nigeria: Reflections on opportunities, options and challenges to legal reforms. *The Liverpool Law Review*, 38(2), 105-134. doi:10.1007/s10991-016-9189-8
- Carugati, C. (2018). The 2017 Facebook saga: a competition, consumer and data protection story. *European Competition and Regulatory Law Review*, 2(1), 4-10. doi:10.21552/core/2018/1/4.
- Crabtree, A., Lodge, T., Colley, J., Greenhalgh, C., Mortier, R., & Haddadi, H. (2016). Enabling the new economic actor: Data protection, the digital economy, and the databox. *Personal and Ubiquitous Computing*, 20(6), 947-957. doi:10.1007/s00779-016-0939-3.
- Cvik, E. D., MacGregor Pelikánová, R., & Malý, M. (2018). Selected issues from the dark side of the general data protection regulation. *Review of Economic Perspectives*, 18(4), 387-407. doi:10.2478/revecp-2018-0020.
- Formánek T. & Tahal R. (2018). Analysis of personal data-sharing consent factors, with focus on loyalty programs in the Czech Republic. *Business: Theory and Practice*, 19, 70-79. doi:10.3846/btp.2018.08.
- Fox, J. (2016, May 30). *polycor: Polychoric and Polyserial Correlations*. R package version 0.7-9. Retrieved from <https://CRAN.R-project.org/package=polycor>.
- Helberger, N., Borgesius, F. Z., & Reyna, A. (2017). The perfect match? A closer look at the relationship between EU consumer law and data protection law. *Common Market Law Review*, 54(5), 1427-1465.
- Hoofnagle, C. J., van der Sloot, B. & Borgesius, F. Z. (2019). The European Union general data protection regulation: what it is and what it means. *Information & Communications Technology Law*, 28(1), 65-98. doi:10.1080/13600834.2019.1573501.
- Hribar, D., Dvojmoč, M., & Markelj, B. (2018, May 30). The impact of the EU General Data Protection Regulation (GDPR) on Mobile Devices. *Varstvoslovje, Journal of Criminal*

- Justice and Security*, 20(4), 414-433. Retrieved from https://www.fvv.um.si/rV/arhiv/2018-4/02_Hribar_Dvojmoc_Marklej_rV_2018-4.pdf.
- Knyrim, R., & Trieb, G. (2011). Smart metering under EU data protection law. *International Data Privacy Law*, 1(2), 121-128. doi:10.1093/idpl/ipr004.
- Koops, B. J. (2014). The trouble with European data protection law. *International Data Privacy Law*, 4(4), 250-261. doi:10.1093/idpl/ipu023.
- Korže, B., & Čertanec, A. (2017). Protecting personal data in the context of interoperability among organizations for protection and rescue. *International Data Privacy Law*, 7(4), 303-313. doi:10.1093/idpl/ix017.
- Kramer, L. C. (2002). Private eyes are watching you: Consumer online privacy protection-lessons from home and abroad. *Texas International Law Journal*, 37(2), 387-420.
- Li, H., Yu, L., & He, W. (2019). The impact of GDPR on global technology development. *Journal of Global Information Technology Management*, 22(1), 1-6. doi:10.1080/1097198X.2019.1569186.
- Lumley, T., Diehr, P., Emerson, S. & Chen, L. (2002). The Importance of the Normality Assumption in Large Public Health Data Sets. *Annual Review of Public Health*, 23(1), 151-169. doi:10.1146/annurev.publhealth.23.100901.140546.
- Martinson, E. O., & Hamdan, M. A. (1971). Maximum likelihood and some other asymptotically efficient estimators of correlation in two way contingency tables. *Journal of Statistical Computation and Simulation*, 1(1), 45-54. doi:10.1080/00949657208810003.
- Moerel, L. (2011). The long arm of EU data protection law: Does the data protection directive apply to processing of personal data of EU citizens by websites worldwide? *International Data Privacy Law*, 1(1), 28-46. doi:10.1093/idpl/ipq004.
- Murphy, M., & Barton, J. (2014). From a sea of data to actionable insights: Big data and what it means for lawyers. *Intellectual Property & Technology Law Journal*, 26(3), 8-17.
- Nyoni, P., & Velempini, M. (2018). Privacy and user awareness on Facebook. *South African Journal of Science*, 114(5/6), 27-31. doi:10.17159/sajs.2018/20170103.
- Pearson, K. (1900). III. Mathematical contributions to the theory of evolution.—VIII. On the inheritance of characters not capable of exact quantitative measurement. — Part I. Introductory. Part II. On the inheritance of coat-colour in horses. Part III. On the inheritance of eye-colour in man. *Philosophical Transactions of the Royal Society of London. Series A*. doi:10.1098/rsta.1900.0024.
- Revelle, W. (2018, May 30). *psych: Procedures for Personality and Psychological Research*. Northwestern University, Evanston, Illinois, USA. Retrieved from <https://CRAN.R-project.org/package=psych>.
- Ruggieri, R., Savastano, M., Scalingi, A., Bala, D., & D'Ascenzo, F. (2018). The impact of Digital Platforms on Business Models: an empirical investigation on innovative start-ups.

- Management & Marketing. Challenges for the Knowledge Society*, 13(4), 1210-1225.
doi: 10.2478/mmcks-2018-0032.
- Saal, F. E., Downey, R. G. & Lahey, M. A. (1980). Rating the ratings: Assessing the psychometric quality of rating data. *Psychological Bulletin*, 88(2), 413-428.
doi:10.1037/0033-2909.88.2.413.
- 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.
doi:10.5210/fm.v19i12.5008.
- Veghes, C., Acatrinei (Pantea), C., Balan, D., & Lalu, B. (2009). European union consumers' views on the protection of their personal data: an exploratory assessment. *Annales Universitatis Apulensis Series Oeconomica*, 11(2), 988-995.
- Veghes, C., Orzan, M., Acatrinei, C., & Dugulan, D. (2012). Privacy literacy: what is and how it can be measured? *Annales Universitatis Apulensis Series Oeconomica*, 14(2), 704-711.
- Voges, K. W. (2014). Improving performance in for-profit contracts: A study of buyer-seller communication and red tape. *Management & Marketing. Challenges for the Knowledge Society*, 9(4), 385-402.
- Wackerly, D., Mendenhall, W., & Scheaffer, R. (2008). *Mathematical Statistics with Applications (7th ed)*. Belmont, CA: Brooks/Cole.
- Woo, J. (2006). The right not to be identified: privacy and anonymity in the interactive media environment. *New Media & Society*, 8(6), 949-967.
doi:10.1177/1461444806069650.
- Woo, J. (2006). "The right not to be identified: privacy and anonymity in the interactive media environment", *New Media and Society*, Vol. 8, No. 6, pp. 949-967.
doi:10.1177/1461444806069650.