The Need for Biometric Anti-spoofing Policies: The Case of Etsy

Mohsen Jozani[1], Gianluca Zanella[2], Maxium Khanov[3], Gokila Dorai[1(✉)], and Esra Akbas[4]

[1] Augusta University, Augusta, USA
{mjozani,gdorai}@augusta.edu
[2] University of Texas at San Antonio, San Antonio, USA
gianluca.zanella@utsa.edu
[3] University of Wisconsin Madison, Madison, USA
mkhanov@wisc.edu
[4] Georgia State University, Atlanta, USA
eakbas1@gsu.edu

Abstract. Effective, safe, and fast identity recognition is crucial in today's rapidly growing society. As a convenient and reliable alternative to traditional identification methods, biometric technologies are increasingly adopted for security applications, such as the verification of ID cards or passports and the authentication of computer and mobile devices. However, if spoofed, such technologies can create serious privacy and security risks, and the proliferation of high quality multimedia content on social media platforms facilitates such spoofing attacks. Unfortunately, many users are unaware of the risks of posting their biometric information online and social media companies are not taking appropriate action to protect them. In this paper, we make the case for biometric anti-spoofing policies by examining the social media enabled marketplace of Etsy. We demonstrate that biometric information can be collected from social media users and that the level of privacy concerns is not a predictor of a user's biometric information sharing behavior.

Keywords: Privacy in Social-media · Digital multimedia forensics · Biometrics

1 Introduction

The term "Biometry" has been used to refer to the field of statistical methods applicable to a wide range of topics in biology. Recently the term Biometrics has also been used to refer to the emerging field of the automated recognition of people based on intrinsic physical or behavioral traits, such as those based on retina-scans, iris-patterns, fingerprints, or face recognition. Facing a steady growth of the population and a "digitalization" of services, many countries around the world have already started making efforts to establish biometric

S. Goel et al. (Eds.): ICDF2C 2022, LNICST 508, pp. 295–306, 2023.
https://doi.org/10.1007/978-3-031-36574-4_17

identity of their citizens. On the other side, an increasing number of private organizations are leveraging biometric identification for applications like employee attendance, door security and logical access. Manufacturers are integrating biometric technologies into mobile and computer devices for fingerprint, face, and iris recognition.

Biometric technologies are becoming widely popular and are replacing the traditional identification methods in our daily computing devices. Entering passwords and PINs, or swiping patterns (which may take several frustrating attempts), are being replaced by fingerprint scanning or using Face ID as a user picks up their phone from the table and looks at their screen. The major advantages of biometric technologies over their traditional counterparts are their convenience, portability, and presumed safety, as they cannot be circumvented by hacked passwords or duplicate ID cards.

However, while the public may perceive biometric identification as a safe and fraud-proof process, biometric technologies such as finger and facial scanning have been shown to be susceptible to spoofing attacks [1]. In fact, it is relatively easy to spoof an off-the-shelf face recognition system using a picture downloaded from social media [2], and a fingerprint replica made of wood glue and printed out with special ink that mimics the conductive properties of human skin can be used to unlock fingerprint-protected devices [3]. These spoofing attacks are often successful, and they are listed as medium level threats in the National Vulnerability Database of the National Institute of Standards and Technology (NIST) in the United States. The fact that leaked biometric information can potentially threaten not just our information security, but also our physical security [4] warrants further study.

Popular online social networks (OSN) constantly encourage users to share content from their daily lives. As a result, photos, voice recordings, details about individuals' jobs, social/family lives, hobbies, and their geolocation data can often be collected from their profiles and triangulated across platforms to reveal sensitive information about them. Besides, there is a growing amount of biometric data and personally identifiable information (PII) present on online platforms [5].

While most users understand the privacy risks associated with posting sensitive textual data, many are not aware of the security threats of posting nontextual data, such as pictures and videos. Such content can be manipulated [6] or misused to impersonate legitimate users [2,4]. On average, users interact with about 7 social media platforms [7], and many think their multiple accounts are not related to each other, without realizing how much sensitive information can be obtained by connecting a single user's profiles across different platforms and merging data from multiple sources. Since each OSN may focus on a specialized need such as socializing, health issues, and professional or academic network, merging data from several sources can provide an accurate presentation of a given user. Moreover, users can be aware of privacy risks associated with their sharing habits without realizing the potential security risks. For example, privacy concerned users may disregard the risks connected to sharing a close-up

picture of their hand because it does not create a privacy threat, but the picture may be used to collect their fingerprints, thus enabling identity theft and other security-related risks. From a theoretical point of view, it would be interesting to test the interplay of privacy-based and security-related behavioral aspects of social media users. It must be established if users that are cautious about exposing their private information on online platforms will still reveal a wide amount of biometric data across their accounts, mainly sharing pictures and videos. Unfortunately, past studies have only focused on leveraging metadata to expose privacy and security potential threats [8].

In this study, we explore the potential threat of collecting biometric information and PII from OSN user profiles. This exploratory research focuses on social media influencers and sellers because they generally post many videos, pictures, and text materials on specific platforms, such as online marketplaces. We argue that the availability of such data may allow malicious actors to collect and merge biometric and other PII data which, in turn, can pose serious privacy and security threats. The goal of this study is to understand whether privacy concerned users are cautious about posting multimedia content that may contain PII and if their proactive behavior is successful in preventing security threats.

We focus on the social media enabled e-commerce website, Etsy, and collect data from each Etsy store, the seller profile associated with the store, and the images and text data of each item sold at the store. Then, we calculate a privacy risk score for each seller and develop a machine learning (ML) based image recognition approach to identify biometric data shared by each seller in the data set.

This research examines the interplay of privacy and security and sheds light on the largely ignored security issue around the multimedia content shared on OSNs. The rest of the paper will present a brief theoretical introduction, the design of this study, the results, and a brief discussion and conclusion section.

2 Theoretical Background

2.1 Privacy

Privacy and privacy concerns play a crucial role on user's online behavior. Active participation in online social communities and networks satisfies people's fundamental needs, such as the need for social relationships [9], social support, self-presentation [10], emotional connection and entertainment [11], identity construction [12], and social capital [13,14]. The perceived benefits of sharing information within a social collective often outweighs the perceived privacy concerns [15]. The decision of sharing information on online platforms is the result of an exchange paradigm process in which the perceived rewards are evaluated against the potential threats to the user's privacy [16]. The result of the (privacy) calculus is often in favor of the leak of private information on online platforms that rewards with social benefits, even in presence of serious threats to the individual privacy. Unfortunately, the body of literature that concerns online privacy does

not take into account security or, to some extent, assumes security as a concept that overlaps privacy.

The definition of privacy refers to the ability of "claiming full control on when, how, and to what extent information about them is communicated to others" [17] while security is referred to as the technological guarantees that ensure that the personal information is transmitted and stored in such a way that third parties are not able to access or tamper with it [18]. Ordinary users often fail to distinguish between security and privacy because they focus exclusively on protecting the availability of personal data on online platforms [19]. This, in turn, prevents users from enacting mitigating strategies for security threats. For example, influencers and sellers on online marketplaces post multimedia content that can trigger online conversations about the brands they endorse [20] that, in turn, increase sales and create product awareness. Given that visual communication is more effective in marketing a product or service, influencer posts often include close-up pictures. These pictures, both face and hands, are the potential source of biometric information that can be then connected to PII information gathered from multiple sources of data.

The technological innovations of the past few years have changed the breadth and depth of potential exposure of private information, mainly because of the increased number of third parties involved in the provision of mobile-enabled services [21]. The growing perception of these risks is reflected in people's increasing concerns about their privacy and the collection and use of their personal information [22]. However, despite of the great concerns about the risks related to privacy leaks, users still expose a great amount of sensitive information across multiple platforms. There are various explanations for this apparent paradox, all based on a mixture of rational and non-rational factor in the human decision-making process. On top of rational cognitive processing, innate limitations such as information asymmetry can make difficult to estimate the potential of the privacy risk. This can be the case of the risks connected to multiple OSN accounts, that are not easily recognized by regular users. Other psychological factors, such as optimism or temporal construal [16] can contribute to over valuate the social rewards from online data leaks of personal data and, at the same time, under valuate the potential risks. This is especially true when the rewards are not only social but also economic, such in the case of social media influencers. These theories explain why people fail to protect personal information even if they are concerned for their privacy. Indeed, past studies find that privacy concerns negatively affect, but do not prevent, online privacy leaks or data exposure. To make things worse, new technologies are creating opportunities to expose biometric information that can be used to access digital and physical private spaces.

2.2 Biometric Identification

Biometrics is a multidisciplinary field concerned with measuring specific biological traits that can be used as an individualized code for recognition. The need and the complexity of identity recognition is increasing because of the population growth and increased mobility. Biometrics is considered as an indispensable

tool to overcome these challenges. While passwords or badges can be easily stolen and used by an intruder, biometric measures have the unique advantage to truly verify that a person is in fact who he claims to be. However, there is an inevitable dilemma in accepting biometrics as private. It is almost impossible to claim that our facial images are private whilst they are captured by surveillance cameras or even shared on social media platforms. Our voices are recorded by most phone-based services or shared through TikTok videos. Therefore, the concern of identity theft prevents the adoption of biometrics as mainstream form of identification in high-security applications [23]. Contrary to password-protected systems, biometric information is widely available and extremely easy to retrieve from websites such as Flickr or Facebook. On the other hand, it can be argued that fingerprint biometrics are more private, in the sense that we don't explicitly share them on social media platforms. In addition, forensic experts have shown limited ability to detect forgeries in the case of fingerprints that are fabricated carefully with well-chosen and processed materials. In fact, spoofing attacks on fingerprint sensors using artificial fingerprint films are successful 80% of the times [3]. As in case of the privacy in social media, the perceived benefits of biometric-based security identification are countered by related risks. Indeed, the perceived security of our fingerprints as authentication method does not trigger as many concerns as other methods do. The perceived security emerges as a critical factor to build user's trust on technology that, in turn, affects the intention to use it and the frequency of usage. However, the assumption of intrinsic security of our fingerprints may prove wrong.

3 Methodology

3.1 Data Collection

To build the dataset for this study, we searched for small handmade items that sellers often photograph holding in their hands. As the pictures of these items are more susceptible to contain fingerprint data. After careful examination of the platform, we decided to focus on five keywords: *flower, keychain, lanyard, pin, and ring*. We built a Python web scraper that first searches for Etsy stores associated with the above keywords and retrieves store-level data such as name, description, location, rating, and the number of items available in each store. Next, since each store is linked to a seller's personal Etsy page, our script collects the seller's name, profile picture, biographic information, location, number of followers, number of following, and items they liked. Finally, the script collects item-level data, including item description, price, shipping method, and all the item pictures posted by the seller. Our dataset contains the data for over 200,000 items sold in 6636 stores.

3.2 Privacy Risk Score

Like users of any other social media platform, Etsy sellers have different tolerance levels for privacy risks and use privacy controls to adjust the type and extent of

information they disclose about themselves [24]. Although some cautious sellers use a nickname (for example, flowergirl79) or their store name on their profile, others use a phrase that includes their first name (for example, Amy's store), and some even disclose their full name on their profile page. Besides, they may decide to post no profile photo, a photo that represents their business (such as a logo or a product portfolio), or a personal picture of themselves. They may also vary in disclosing their location, biographic information, and social and platform interactions.

To understand sellers' privacy preferences, we first use a BERT-based named entity recognition model to examine if they disclose their first and last names [25]. Then, we use a machine learning approach by OpenCV [26] to determine if the seller has posted an identifiable photo of themselves on their profile page. Overall, we extract seven privacy items from each seller's profile page: *first_name, last_name, location, like, follow, picture, biography.*

Based on these items, we compute a total privacy risk score for each seller following Liu and Terzi [27] naïve privacy score computation framework. For N number of sellers and n number of privacy items, we define a matrix with size $N \times n$ where the range of items i is $1 \leq i \leq n$ and the range of sellers j is $1 \leq j \leq N$.

Every item for each seller takes a binary label (0 or 1). If seller j disclosed information regarding item i, that item takes the response value of 1 *($R_{i,j}$ = 1)*. Otherwise, if the seller did not disclose or made that information private, *($R_{i,j}$ = 0)*.

Privacy risk score is measured using the two dimensions of sensitivity and visibility of information.

(a) Sensitivity

Sensitivity is measured across all users for each privacy item and describes the general likelihood of publicly sharing a specific piece of information. That is:

$$S_i = \frac{N - R_i}{N} \tag{1}$$

where R_i is the sum of all (non-zero) instances of i. Some items are naturally more sensitive than others (such as follow, biography, and picture) and therefore, sellers are less likely to disclose them. Table 1 shows the sensitivity scores calculated for all seven privacy items.

(b) Visibility

Visibility of a privacy item i depends on its value across the entire sample, as well as the perception and valuation of user j. It is computed as:

$$V_{i,j} = \frac{R_i}{N} \times \frac{R_j}{n} \tag{2}$$

The higher the value of V_i, the less sensitive is the item.

Table 1. Sensitivity scores

No	Item	Sensitivity Score
1	Follow	0.707
2	Biography	0.657
3	Picture	0.649
4	Last Name	0.603
5	Like	0.469
6	Location	0.257
7	First Name	0.248

The total privacy risk score for user j is, therefore, the sum of the product of j's sensitivity and visibility scores for each privacy item i:

$$PR\_Score_j = \sum_i S_i \times V_{i,j} \tag{3}$$

Figure 1 shows privacy risk score distribution for all sellers. The minimum value of zero represents sellers who disclosed none of the privacy items and the highest disclosure (privacy leak) value for sellers who disclosed all seven items is 1.31. The sample has a normal distribution with mean and median values of 0.67 and 0.65, respectively.

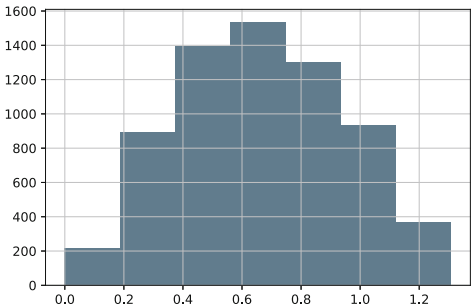

Fig. 1. Distribution of Privacy risk scores

3.3 Extracting Fingerprint Data

To find images containing fingers in the collected dataset, we employ a hierarchical finger detection algorithm including 2 steps: hand detection(1) and finger detection (2). In the hand detection step, our primary goal is to eliminate images without hands efficiently. We use a neural network model as a machine learning algorithm for this goal. We utilize the YOLO (V3) - based NN model designed by Alam et al. [28]. The architecture of this model consists of 106 convolutional

Fig. 2. Examples of Good and Bad Finger Identification

layers. The first 53 convolutional layers form a base neural network called Darknet. Darknet was pre-trained on the image net dataset and served as a feature extractor for the network. The subsequent 53 convolutional layers detect the object in the image; in our case, just the part that detects the hand is used. The initial data reduction step runs efficiently with an average speed of 330 images per second. After eliminating many images without a hand, we apply pre-trained Google's MediaPipe framework [29], as a more robust hand and finger detection model but computationally more expensive model. It identifies each hand and joint more accurately with an average speed of 116 images per second. The Mediapipe hand detection framework crops the hands and runs a joint detection model, which consists of a feature extractor that generates all the hand joint positions.

The entire work of extracting fingerprint data is done without using GPUs. Instead, we employ the Ray Python library to parallelize the data analysis. In the first step, the confidence threshold for the hand detection model is intentionally set low to reduce the number of false negatives to preserve the most data. The MediaPipe model used in the second step performs poorly with images like the following: (a) with partially occluded hands; (b) the hand covered the whole image; (c) images that are not related to our case, like toys with fingers, as shown in Fig. 2. Using our hierarchical model, we can flag 2% of the images as containing fingers. We then process the finger data to determine whether any of the sellers had any fingers in their listings.

4 Findings

We performed simple statistical analyses to examine the role of privacy risk on disclosing fingerprint data. From 6,492 seller profiles we examined, 46% posted at least one photo with visible fingerprints. Moreover, we created a binary variable based on the median of privacy risk score ($Q2 = 0.65$) and compared *LOW_PR* vs *HIGH_PR* sellers in terms of fingerprint data leak. While the probability of fingerprint leak for *HIGH_PR* sellers is 48%, *LOW_PR* sellers have 52% probability of leaking their fingerprint data. We also performed logistic regression to examine if privacy risk score can predict fingerprint data leakage. Our finding suggests that privacy risk is not a significant predictor of disclosing fingerprint data ($\beta = -0.0108$, $p = 0.74$).

Our findings suggest that regardless of privacy risk level, people on social media are likely to post their sensitive biometric data and it is up to the platforms to take appropriate measures to protect the security and privacy of social media users.

5 Discussion and Conclusion

This paper focuses on the divide between user protective strategies to mitigate privacy-related risks and security threats. By analyzing data from Etsy users' feeds, we demonstrate that the level of users' privacy concerns does not predict the amount of biometric information they may inadvertently disclose in their social media posts. While privacy concerned users may take strategies such as using nicknames or removing profile photos to mitigate privacy-related risks, they may still inadvertently disclose their biometric information in the images and videos they share. A possible takeaway is that users may not be aware of the potential security threats of posting their biometrics. Perhaps, given the fast pace of technological innovation, it is not reasonable to ask users to be competent and up-to-date in cyber-focused technicalities. Policies and regulations should be in place to require social media platforms to restrict and protect posts with users' PII as they already do with offensive or indecent content. Indeed, our paper joins the many calls for privacy policies that take into account the real dispersion and depth of sensitive information [30].

5.1 Research Implications

This research contributes to the growing body of research on the ethical implications of the use of biometrics for identifying and authenticating people. On one hand, the use of biometrics raises difficult questions regarding data protection. More directly theoretical questions concern the conceptualisation of persons as a "machine-readable body" [31], as well as the role of biometrics in various sociopolitical and economic settings [32]. A part of the ethical issue that biometrics information changes overtime in aging individuals [33], biometrics poses a

serious threat to individual security because it falls outside the radar of an individual's privacy concerns. This calls for the evolution of the concepts of privacy as a multi-factor or multi-domain concept that should include also security-related components.

5.2 Practical Considerations

It can be observed that users are increasingly dissatisfied with the policies of online social media companies. At the same time, policymakers are trailing behind technical innovation, with regulations that are not applicable to up-to-date technologies. The European general data protection regulation (GDPR) is considered a milestone in this sense, because it leaves flexibility for technological advancements [34]. The GDPR specifically recognizes biometric data as a subset of sensitive personal data deemed a "sensitive category of personal data." Still, it does not explicitly consider the case of biometric data casually embedded in shared pictures. Until the legal ramifications of this gray area are clarified, the first approach is to call social media companies into action. As companies tag posts as "not verified" or offensive, they should alert the users when a post may contain biometric data. Offering this feature will improve not only the user's security, but will also contribute in building trust between customers and companies, which will benefit both.

References

1. Ratha, N.K., Connell, J.H., Bolle, R.M.: Enhancing security and privacy in biometrics-based authentication systems. IBM Syst. J. **40**(3), 614–634 (2001)
2. Wen, D., Han, H., Jain, A.K.: Face spoof detection with image distortion analysis. IEEE Trans. Inf. Forensics Secur. **10**(4), 746–761 (2015)
3. Mott, N.: Hacking fingerprints is actually pretty easy-and cheap, November 2021
4. Alnabhi, H., Al-naamani, Y., Al-madhehagi, M., Alhamzi, M.: Enhanced security methods of door locking based fingerprint. Int. J. Innov. Technol. Explor. Eng. **9**(03), 1173–1178 (2020)
5. Girelli, C.M.A., et al.: Application of a standard procedure to avoid errors when comparing fingerprints with their reversals in fake documents. J. Forensic Sci. Med. **2**(1), 60 (2016)
6. Boididou, C., et al.: Verifying information with multimedia content on twitter. Multimed. Tools Appl. **77**(12), 15545–15571 (2018)
7. Dean, B.: Social network usage & growth statistics: how many people use social media in 2021, vol. 2, p. 2021 (2021). Accessed July 2021
8. Gouert, C., Tsoutsos, N.G.: Dirty metadata: understanding a threat to online privacy. IEEE Secur. Priv. **01**, 2–9 (2022)
9. Krämer, N.C., Schäwel, J.: Mastering the challenge of balancing self-disclosure and privacy in social media. Curr. Opin. Psychol. **31**, 67–71 (2020)
10. Kim, J., Tussyadiah, I.P.: Social networking and social support in tourism experience: the moderating role of online self-presentation strategies. J. Travel Tour. Mark. **30**(1–2), 78–92 (2013)

11. Sheth, S., Kim, J.: Social media marketing: the effect of information sharing, entertainment, emotional connection and peer pressure on the attitude and purchase intentions. GSTF J. Bus. Rev. (GBR) **5**(1) (2017)
12. Berger, C.R., Calabrese, R.J.: Some explorations in initial interaction and beyond: toward a developmental theory of interpersonal communication. Hum. Commun. Res. **1**(2), 99–112 (1974)
13. Ellison, N.B., Steinfield, C., Lampe, C.: The benefits of Facebook "friends:" social capital and college students' use of online social network sites. J. Comput.-Mediat. Commun. **12**(4), 1143–1168 (2007)
14. Ellison, N.B., Steinfield, C., Lampe, C.: Connection strategies: social capital implications of Facebook-enabled communication practices. New Media Soc. **13**(6), 873–892 (2011)
15. Kokolakis, S.: Privacy attitudes and privacy behaviour: a review of current research on the privacy paradox phenomenon. Comput. Secur. **64**, 122–134 (2017)
16. Hallam, C., Zanella, G.: Online self-disclosure: the privacy paradox explained as a temporally discounted balance between concerns and rewards. Comput. Hum. Behav. **68**, 217–227 (2017)
17. Westin, A.F.: Privacy and freedom. Wash. Lee Law Rev. **25**(1), 166 (1968)
18. Mekovec, R., Hutinski, Ž.: The role of perceived privacy and perceived security in online market. In: 2012 Proceedings of the 35th International Convention MIPRO, pp. 1549–1554. IEEE (2012)
19. Flavián, C., Guinalíu, M.: Consumer trust, perceived security and privacy policy: three basic elements of loyalty to a web site. Ind. Manag. Data Syst. (2006)
20. De Veirman, M., Cauberghe, V., Hudders, L.: Marketing through instagram influencers: the impact of number of followers and product divergence on brand attitude. Int. J. Advert. **36**(5), 798–828 (2017)
21. Jozani, M., Ayaburi, E., Ko, M., Choo, K.-K.R.: Privacy concerns and benefits of engagement with social media-enabled apps: a privacy calculus perspective. Comput. Hum. Behav. **107**, 106260 (2020)
22. Madden, M.: Public perceptions of privacy and security in the post-snowden era, (2014)
23. Schuckers, S.A.: Spoofing and anti-spoofing measures. Inf. Secur. Tech. Rep. **7**(4), 56–62 (2002)
24. Cavusoglu, H., Phan, T.Q., Cavusoglu, H., Airoldi, E.M.: Assessing the impact of granular privacy controls on content sharing and disclosure on Facebook. Inf. Syst. Res. **27**(4), 848–879 (2016)
25. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.,: Bert: pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018)
26. Bradski, G., Kaehler, A.: Opencv. Dr. Dobb's J. Softw. Tools **3**, 2 (2000)
27. Liu, K., Terzi, E.: A framework for computing the privacy scores of users in online social networks. ACM Trans. Knowl. Discov. Data (TKDD) **5**(1), 1–30 (2010)
28. Alam, M.M., Islam, M.T., Rahman, S.M.: Unified learning approach for egocentric hand gesture recognition and fingertip detection. Pattern Recogn. **121**, 108200 (2022)
29. Lugaresi, C., et al. : Mediapipe: a framework for perceiving and processing reality. In: Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR) 2019 (2019)
30. Patsakis, C., Zigomitros, A., Papageorgiou, A., Galván-López, E.: Distributing privacy policies over multimedia content across multiple online social networks. Comput. Netw. **75**, 531–543 (2014)

31. Ploeg, I.V.D., Lyon, D.: Biometrics and the body as information: normative issues of the socio-technical coding of the body. Surveillance as Social Sorting: Privacy, Risk, and Digital Discrimination. Londres e Nova Iorque, Routledge, pp. 57–73 (2002)
32. Agamben, G., Murray, S.J.: No to biopolitical tattooing. Commun. Crit. Cult. Stud. **5**(2), 201–202 (2008)
33. Rebera, A.P., Mordini, E.: Biometrics and ageing: social and ethical considerations. Age Factors Biom. Process. 37–58 (2013)
34. Goddard, M.: The EU general data protection regulation (GDPR): European regulation that has a global impact. Int. J. Mark. Res. **59**(6), 703–705 (2017)