Should User-generated Content be a Matter of Privacy Awareness?
A Position Paper

Nicolás Emilio Díaz Ferreyra, Rene Meis and Maritta Heisel
RTG User-Centred Social Media, University of Duisburg-Essen, Germany

Keywords: Adaptive Privacy, Self-disclosure, Awareness, Social Network Sites, Data Visceralization.

Abstract: Social Network Sites (SNSs) like Facebook or Twitter have radically redefined the mechanisms for social interaction. One of the main aspects of these platforms are their information sharing features which allow user-generated content to reach wide and diverse audiences within a few seconds. Whereas the spectrum of shared content is large and varied, it can nevertheless include private and sensitive information. Such content of sensitive nature can derive in unwanted incidents for the users (such as reputation damage, job loss, or harassment) when reaching unintended audiences. In this paper, we analyse and discuss the privacy risks of information disclosure in SNSs from a user-centred perspective. We argue that this is a problem of lack of awareness which is grounded in an emotional detachment between the users and their digital data. In line with this, we will discuss preventative technologies for raising awareness and approaches for building a stronger connection between the users and their private information. Likewise, we encourage the inclusion of awareness mechanisms for providing better insights on the privacy policies of SNSs.

1 INTRODUCTION

In 1966, tobacco companies across the United States were affected by a law that later on changed the standards for the commercialization and distribution of cigarettes. For the first time in the history, a legislation requiring warnings about the risks associated with the consumption of tobacco was proposed by the U.S Congress (Hiilamo et al., 2014). Since then, the companies began fighting against Health Warning Labels (HWLs) in cigarette packs basically arguing that people already knew the hazards of smoking. Despite their efforts on blocking or weaken HWLs, nowadays many countries have included and implemented HWL in their legislations (Hiilamo et al., 2014).

Social Network Sites (SNSs) are spaces which are not free of privacy risks, and like in the case of cigarettes consumers, users of SNSs might have heard about some of these risks before or during their activity period (i.e. before or after opening an account in a SNS). While one might argue that the risks of disclosing personal or sensitive information in SNSs are not as severe as the risks of smoking, unwanted incidents such as job loss, reputation damage, or unjustified discrimination should not be neglected or disregarded. However, very little information (for not to say none) is provided by the SNSs about the potential risks of information sharing.

Privacy policies can be considered as an initial approach towards the information on potential privacy risks. However, these electronic documents are shown once to the users (when registering), and are hardly revised by them in the future. Moreover, privacy policies basically inform about which data is collected, how is processed, and under which conditions it is disclosed to third parties; without any emphasis on informing about potential risks. If we add to this that users are not strongly attached to their private information, then the chances of users regretting to have shared private information increases.

We believe that, like tobacco consumers, the users of SNSs should be empowered with information about the potential risks of information sharing. Moreover, we believe that awareness mechanisms can be a good alternative not only to inform the users about such risks, but also to create a stronger tie between them and their private information. In this paper we take a closer look at the privacy risks associated with user-generated content in SNSs in order to discuss possible solutions to this issue. Moreover, we provide arguments towards the use of adaptive preventative technologies to move towards a more

DÃaz Ferreyra N., Meis R. and Heisel M.
Should User-generated Content be a Matter of Privacy Awareness? - A Position Paper.
DOI: 10.5220/0006517302120216
In *Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (KMIS 2017)*, pages 212-216
ISBN: 978-989-758-273-8

privacy-aware social environment in SNSs.

The rest of the paper is organized as follows. In the next section, we discuss the motivation scenario and the paper's background. In Section 3, we analyse the role of privacy policies and media technologies on modulating users' perceptions towards their private data. Following, we discuss in Section 4 preventative technologies for the generation of awareness within SNSs. In Section 5, we analyse an approach for incorporating privacy heuristics derived from users' regrettable experiences into the design of preventative technologies. Thereafter, we discuss the advantages and drawbacks of this approach in Section 6. Finally, we conclude in Section 7 with an outlook and considerations for further research.

2 MOTIVATION AND BACKGROUND

In 2018, the EU's new General Data Protection Regulation (GDPR) (Regulation, 2016) will come into force as the conclusion of a hectic debate which has involved academics, Internet service providers and international organizations across the world. For many, the Internet is considered an open platform for democratic participation which promotes freedom of expression and the right to information access. Therefore, is not surprising that the GDPR, and more specifically the Right to be Forgotten (RTBF), raises concerns related to abusive removal demands of user-generated content (e.g. public officers trying to suppress criminal records), and other issues about potential unjustified censorship[1]. This is a debate which mainly circles around the right to erase or de-list information put online by another Internet user. In this work, we do not aim to discuss this aspect of the GDPR. Instead, we look to resume the discussion to the information that users disclose about themselves in SNSs.

One of the critical concepts included in the GDPR is the one of "personal data". For instance, Article 4 says that information related to a "data subject" (i.e. an identified or identifiable natural person[2]), such as name, identification number, location, factors specific to his/her physical, physiological, genetic, mental, economic, cultural or social identity, should be considered as personal information and therefore require *unambiguous* consent to be processed. Likewise, Article 9 says that racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade-union membership should not be processed unless the data subject gives *explicit* consent. *Unambiguous* consent can be given through a conduct that clearly indicates that the data subject agrees with the proposed processing of his/her personal data (e.g. when telling the doctor about the medical ailment one is suffering while he/she enters notes in a computer system). On the other hand, *explicit* consent should be given trough an explicit action by the data subject. This is normally granted after the data subject clicks on 'Yes, I agree" in the *privacy policies* of the service provider.

3 DATA "VISCERALIZATION"

One of the main reasons for differentiating private data from general data are the risks associated with their inappropriate processing and public disclosure. Basically, the GDPR encloses an implicit warning message for data processors (the SNSs in our case) which is that they should safeguard data subjects from unwanted incidents (such as unjustified discrimination, political or religious persecution, or fraud) by treating carefully their personal data. Privacy policies, on the other hand, inform the users about which information will be collected, processed, used, disclosed and managed by the data processor. As one can observe, there is a semantic difference in the message of the GDPR and the one of privacy policies. Whereas the GDPR endows service providers with a better perception on the importance of the users' personal information, privacy policies do not provide cues to data subjects about the importance of their own personal information. Consequently, privacy policies in some point modulate users' perceived severity of privacy risks in SNSs.

Like privacy policies, information sharing interfaces of SNSs also play an important role in shaping our perceptions of information privacy (Stark, 2016; Díaz Ferreyra et al., 2017a). Such interfaces are the entry point of user-generated content which, in many cases, contains private information. However, since digital data is intangible, information sharing interfaces of SNSs regulate users' emotional perception and attachment towards their private information. Let us consider the following example: imagine that a stranger stops you in the street and asks you for your passport. It is quite unlikely that someone would grant this request in the real world. Moreover, this situation would normally come along with a *visceral reaction* (i.e. an instinctive gut-deep bodily response

[1] The work by Keller (Keller, 2017), offers a clarifying view on the RTBF and its hazards for freedom of expression and information rights on the Internet.

[2] An identifiable person is one who can be identified, directly or indirectly.

like a burning sensation in the stomach) as consequence of this unexpected request. However, when this information is requested through the interfaces of a SNSs, such reactions do not seem to arise. Consequently, privacy policies and sharing interfaces are not succeeding in taking the users' emotional perception of their private data to the visceral level.

4 PRIVACY AWARENESS IN SNSs

Like in the case of HWL for the commercialization of cigarettes, awareness mechanisms for SNSs can contribute to bridge the emotional gap between users and their digital data. In this section we discuss different preventative technologies oriented to generate awareness in online self-disclosure scenarios. This is, scenarios in which users intend to reveal their own private information in SNSs.

4.1 Preventative Technologies

Different preventative technologies have been proposed for mitigating the unwanted consequences of online self-disclosure (Calikli et al., 2016; Díaz Ferreyra et al., 2016; Fang and LeFevre, 2010; Wang et al., 2013; Ghazinour et al., 2013). One of the most representative of these approaches is the one by Wang et al. (Wang et al., 2013) consisting of three plugins for Facebook. These plugins called "privacy nudges" intervened when the user was about to post a message in his/her Facebook biography either (i) introducing a delay, (ii) providing visual cues about the audience of the post, or (iii) giving feedback about the meaning (positive or negative) of the post. However, since the feedback generated by the nudges was the same for every user of Facebook, they did not succeed on reaching high levels of acceptance. This is, some users liked them and others found them annoying. Consequently, this type of technology should provide adaptive feedback and awareness to their users in order to being widely adopted.

4.2 Adaptive Awareness

Adaptive preventative technologies seek to develop mechanisms capable to provide tailored feedback and awareness to their users. One of the preventative technologies which follow this direction is the one of Ziegeldorf et al., consisting in a framework of personalized privacy metrics for the generation of adaptive awareness (Ziegeldorf et al., 2015). This approach, called Comparison-based Privacy (CbP), consists of analysing different *comparison metrics* which are computed over the content being shared among different *comparison groups*. Basically, comparison groups consist of groups of people with which the user can intuitively relate to (e.g. family, friends and colleagues, users with the same profession or same age). Likewise, comparison metrics capture aspects of the sharing behaviour within a privacy group, such as the sentiment and the type of the content being shared. A user can choose for instance to compare the amount of hate speech in his/her posts against the one of people with his/her same profession. If this value exceeds a given threshold, then the system alerts the user. Thresholds can be set individually by users, or according to general profiles representing different privacy attitudes (e.g. unconcerned, pragmatist or fundamentalist). Approaches like this one overcome the engagement issue caused by generic warning messages of static approaches.

5 VISCERAL-AWARENESS DESIGN

One of the key elements of HWLs in cigarettes packaging is that they include pictorial representations of the risks of tobacco consumption. This is done in order to make users perceive such risks in a more visceral way. In the case of online self-disclosure, regrettable experiences come along with visceral reactions from the users. This is, when a user lives an unwanted incident after disclosing personal data in SNSs, then a feeling of regret and repentance arises together with a visceral reaction. In this section, we discuss design principles introduced by Díaz Ferreyra et al. to include regrettable experiences into the design process of preventative technologies.

5.1 Privacy Heuristics

Díaz Ferreyra et al. propose to take into account regrettable self-disclosure experiences in order to endow preventative technologies with visceral-awareness principles (Díaz Ferreyra et al., 2017a). Basically, they suggest that privacy heuristics (best practices) can be derived from regrettable self-disclosure experiences and used thereafter to raise privacy awareness. For this, they introduce a Privacy Heuristics Derivation Method (PHeDer) for eliciting privacy best practices from user's regrettable experiences (Díaz Ferreyra et al., 2017a). The first step of this method, called *Regret Acknowledgement*, consists on gathering evidence about a regrettable experience and describe it in terms of: (i) the information that was disclosed (ii) the unintended audience it reached, and (iii) the unwanted incidents that lead the user to a feeling of regret. The output of this step can be represented as in Fig. 1, where the user reported to have shared his/her political affiliation in a public post. Once the regrettable scenario is described, it is forwarded to the next step called *Regret Analysis*.

The *Regret Analysis* step consists of refining the scenario of Fig. 1 into privacy risks consisting of a 7-tuple of elements: a list of *personal attributes*, the *unintended audience*, the *unwanted incident*, the *frequency* of the unwanted incident, the *impact* of the unwanted incident, the *risk level*, and the user's *privacy attitude*. Whereas the personal attributes can be derived from articles 1 and 9 of the GDPR, frequency and impact of the unwanted incident (and consequently the risk level) can be expressed using nominal scales. Privacy attitudes are one of the *pragmatist* (medium privacy concern), *fundamentalist* (high privacy concern), or *unconcerned* (low privacy concern). For the example of Fig. 1, the output of this step would be risk([political_opinion], work_colleagues, wakeup_call, likely, major, very_high, pragmatist). This information if then forwarded to the third step of the method which is *Heuristic Design*.

USER'S POST

"Seriously? Trump became president? What is happening to the world!? #republicanssuck"

Actual Audience: PUBLIC.
Unintended Audience: The user's work colleagues.
Unwanted Incidents: Wake-up call from superior.

Figure 1: Example of self-disclosure scenario.

The *Heuristic Design* step uses Constraint Based Modeling (CBM)(Mitrovic and Ohlsson, 2006) as the design principle for encoding the outcome of step 2 into a privacy heuristic. In CBM a constraint consists of a pair of *relevance* and *satisfaction* conditions, where each member of the pair can be seen as a set of features or properties that a disclosure scenario must satisfy. For the given example, the relevance conditions would be the existence of a political opinion inside a post, and the satisfaction condition would be not include the work colleagues in the post audience. Such constraints can be expressed using Horn clauses in Prolog and be included in the final step (*Constraint Integration*) in the Privacy Heuristics Data Base (PHDB) of an Instructional Awareness System (IAS) (Díaz Ferreyra et al., 2017a; Díaz Ferreyra et al., 2016).

5.2 Instructional Awareness

Basically, an IAS uses the heuristics inside a PHDB to detect potentially regrettable disclosures. Such heuristics are evaluated when a "post" event takes place in order to determine if the disclosure action can derive in a regrettable scenario for the user. This is, done first by evaluating the *relevance* condition of the heuristics. Let us consider a scenario where a user wants to disclose once again his/her *political affiliation* inside a public post. Let us also consider that the heuristic discussed in Section 5.1 is part of the PHDB. In this case, IAS will detect that the disclosure can lead to a potential regret, and therefore proceeds to raise a warning to the user.

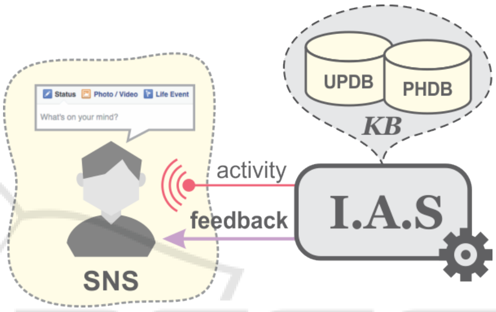

Figure 2: Instructional Awareness System (IAS).

In order to generate adaptive feedback, IAS takes into account adaptivity variables such as the user's privacy attitude, the number of times the user has ignored/accepted the warnings, and how often he/she discloses private information. This information is stored in a User Performance Data Base (UPDB) that, together with the PHDB, makes up IAS's Knowledge Base (KB). With the information stored in its KB, IAS can generate a message such as "Revealing your political affiliation to your work colleagues can bring you problems. Do you want some hints on how to protect your private data?" and recommend the user to restrict the post's audience (for instance to "friends only"). Since information about the risks is also kept in the KB, an IAS can also provide such additional information in the warning message.

6 DISCUSSION

Following a similar approach to the HWLs in cigarettes packages, Díaz Ferreyra et al. propose to inform the users of SNSs about the risks of online self-disclosure though an IAS. For this, IAS requires risk knowledge which is stored in a PHDB and obtained through a privacy heuristics derivation method.

This method is an *offline* approach for eliciting privacy heuristics from regrettable online self-disclosure experiences. Basically, the input of the method are the experiences that users have reported themselves (to the development team of IAS for instance), or the outcome of an empirical research (e.g. questionnaires or face to face interviews). This approach is effective for building a baseline of heuristics prior to the execution of the system. However, eliciting new entries of the PHDB requires the execution of this process which can be expensive and inefficient in terms of the resources and time needed to conduct interviews and process the outcome of them. One way to overcome this issue is to consider deleted posts with private information as potential sources of heuristics (Díaz Ferreyra et al., 2017b). This is, to use such posts as the input of a machine learning engine for the automatic derivation of privacy heuristics at runtime. This way, the PHDB can be updated with new heuristics without having to execute offline iterations of the PHeDer method.

7 OUTLOOK AND CONCLUSION

Adaptive awareness technologies seem to be promising approaches for empowering the users of SNSs in making wiser and more informed decisions, as to protect them from the risks of over-sharing private information. We believe that this is not a minor issue that should be taken seriously into consideration by Internet service providers, multilateral organizations and policy makers. We have used the example of HWLs in cigarettes packages as a motivating scenario for working towards a more privacy aware social environment in SNSs. Certainly, this topic will be part of a in-depth and intense debate in the future. Therefore we hope this paper will offer a more clarifying view on this issue and serve as an instrument for the development of more effective solutions.

ACKNOWLEDGEMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) under grant No. GRK 2167, Research Training Group ”User-Centred Social Media”.

REFERENCES

Calikli, G., Law, M., Bandara, A. K., Russo, A., Dickens, L., Price, B. A., Stuart, A., Levine, M., and Nuseibeh, B. (2016). Privacy Dynamics: Learning Privacy Norms for Social Software. In *Proceedings of the 11th International Symposium on Software Engineering for Adaptive and Self-Managing Systems*, pages 47–56. ACM.

Díaz Ferreyra, N. E., Meis, R., and Heisel, M. (2017a). Online Self-disclosure: From Users' Regrets to Instructional Awareness. In *Proceedings of the IFIP International Cross-Domain Conference (CD-MAKE)*. Accapted for publication.

Díaz Ferreyra, N. E., Meis, R., and Heisel, M. (2017b). Towards an ILP Approach for Learning Privacy Heuristics From Users' Regrets. In *Proceedings of the 4th European Network Intelligence Conference (ENIC)*. Accapted for publication.

Díaz Ferreyra, N. E., Schäwel, J., Heisel, M., and Meske, C. (2016). Addressing Self-disclosure in Social Media: An Instructional Awareness Approach. In *Proceedings of the 2nd ACS/IEEE International Workshop on Online Social Networks Technologies (OSNT)*. ACS/IEEE.

Fang, L. and LeFevre, K. (2010). Privacy wizards for social networking sites. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 351–360, New York, NY, USA. ACM.

Ghazinour, K., Matwin, S., and Sokolova, M. (2013). YourPrivacyProtector: A Recommender System for Privacy Settings in Social Networks. *International Journal of Security, Privacy and Trust Management (IJSPTM)*, 2(4).

Hiilamo, H., Crosbie, E., and Glantz, S. A. (2014). The evolution of health warning labels on cigarette packs: the role of precedents, and tobacco industry strategies to block diffusion. *Tobacco control*, 23(1):e2–e2.

Keller, D. (2017). The right tools: Europe's intermediary liability laws and the 2016 general data protection regulation. Technical report, Stanford Law School Center for Internet and Society.

Mitrovic, A. and Ohlsson, S. (2006). Constraint-based knowledge representation for individualized instruction. *Computer Science and Information Systems (ComSIS) Journal*, 13(1).

Regulation, G. D. P. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union (OJ)*, 59:1–88.

Stark, L. (2016). The Emotional Context of Information Privacy. *The Information Society*, 32(1):14–27.

Wang, Y., Leon, P. G., Scott, K., Chen, X., Acquisti, A., and Cranor, L. F. (2013). Privacy Nudges for Social Media: An Exploratory Facebook Study. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 763–770. ACM.

Ziegeldorf, J. H., Henze, M., Hummen, R., and Wehrle, K. (2015). Comparison-based privacy: nudging privacy in social media (position paper). In *International Workshop on Data Privacy Management*, pages 226–234. Springer.