TU Delft

Delft University of Technology

The Right of Access as a Tool for Privacy Governance

Findings from Individual Requests & Proposal for a Crowd-sourced Dataset of Privacy Practices

Asghari, Hadi; Mahieu, René; Mittal, Prateek; Greenstadt, Rachel

Publication date
2017

Document Version
Final published version

Published in
Proceedings of 17th Privacy Enhancing Technologies Symposium

Citation (APA)
Asghari, H., Mahieu, R., Mittal, P., & Greenstadt, R. (2017). The Right of Access as a Tool for Privacy Governance: Findings from Individual Requests & Proposal for a Crowd-sourced Dataset of Privacy Practices. In *Proceedings of 17th Privacy Enhancing Technologies Symposium* (pp. 1-2)

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

The Right of Access as a tool for Privacy Governance

Findings from Individual Requests & Proposal for a Crowd-sourced Dataset of Privacy Practices

Hadi Asghari
Delft University of Technology
h.asghari@tudelft.nl

René L.P. Mahieu
Delft University of Technology
r.l.p.mahieu@tudelft.nl

Prateek Mittal
Princeton University
pmittal@princeton.edu

Rachel Greenstadt
Drexel University
rachel.a.greenstadt@drexel.edu

1. INTRODUCTION

Personal data is a key asset in the data economy. Ubiquitous internet connectivity, increased computing power, and cheaper storage have resulted in the exponential use of personal data. This has many potential commercial and research advantages [1]. It also has many citizens and scholars worried [2].

In Europe, a leading principle of governance with respect to personal data is that the data subject be able to exert control over the conditions under which their data is collected and used. Since a precondition for exerting control is transparency, European lawmakers created the so called 'right of access', in article 12 of the Data Protection Directive [3]. This right permits a citizen to ask organizations for "confirmation as to whether or not data relating to him are being processed, and information at least as to the purposes of the processing, the categories of data concerned, and the recipients ... to whom the data are disclosed". The right has been retained in the upcoming General Data Protection Regulation [4].

In this talk we present our findings from submitting data access requests to 32 private and public organizations in the Netherlands. Responses from the organizations have been heterogeneous, ranging from evasion tactics, to personal data with varying degrees of detail and completeness. The responses provide a glimpse into what one might describe as disorganized processes. It appears that despite the right of access having existed for almost two decades and exercised by activists in a number of high profile cases (such as Schrems' "Europe vs Facebook" [5]), the majority of organizations lack an internal process for handling them. In fact, two organizations explicitly told to us that our requests were the first they had *ever* received. Still, the right provides some visibility into the types of data organizations collect in addition to how well they manage personal data.

We propose that the right of access can be leveraged systematically to uncover more than individual anecdotal findings, by building a platform to crowd-source and crowd-interpret access requests. The centralized storing, interpretation, and encoding of access requests (by volunteers with privacy considerations) allows us to create a catalog of data collection, processing, and sharing practices in the broader data ecosystem. This dataset may be useful for other PETS researchers. It may also allow a new form of privacy governance: the high-level mapping of privacy practices allows discussions and serves as an external checks and balances (or benchmark). We would like to hear the communities thoughts on these points.

2. INITIAL FINDINGS

Table 1 presents an overview of the results from our access requests. We have submitted the requests on behalf of the first two authors to both public and private entities. The requests were written in Dutch. We sent requests in both letter and email form, to which organizations responded equally.

Table 1. Access Request Responses

Access Requests	No Response	No Data	Some Data	Wrong Data
32 *unique organiz.*	4 *after 3rd reminder*	12 *'deleted' & referral*	15 *actual & categories*	1

Who Responds. The organizational unit signing the responses differed largely. This included legal teams, security teams, privacy officers, and consumer complaint departments.

Response Time. The Dutch law sets a four week time limit for responses. However, the majority of organizations did not respond in this time frame, after which we sent two reminders. Most organizations responded replied after the first reminder – in approximately 6-8 weeks. One entity who has not yet responded explained that they are facing a large backlog of consumer complaints (unrelated to access requests).

Authentication of Requests. Most organizations verified our identity solely on the basis of the copy of the passports, and did not verify the mailing or email address used. About ten organizations however used additional authentication, for instance calling us to verify the request, or in two cases, asking us to visit them in person with our passport. For privacy and security reasons these additional authentication steps are advisable, even though they create additional costs.

Responses with No Data. Twelve organizations responded by saying either they have either deleted the data, or that they do not "control" the data. *Data controller* is a legal term about the entity that holds the data and is responsible for it. We were often given a referral, for instance Master Card referred us to our bank. In some cases however the referral did not make sense. An interesting example is Amsterdam Schiphol airport who stated that the airport holds no personal data and all data is held by airlines and other third

parties[1]. Open requesting clarification about who handles boarding passes and luggage, the airport responded that they actually do process personal data, but that such data is deleted every few days. This may be a great privacy practice, but casts doubt on the veracity of their first answer.

Responses with Data. About half of the fifteen organizations that responded sent back personal data, and the other half described the categories of data (labels) but didn't provide actual data. In some cases, we were informed that we could request additional data, but we might be charged for administrative costs. For example, T-Mobile offered ten days of CDR and cell tower location for a period of our choice. Data were typically printed out, even in response to emails, or a PDF of a print-out was sent. Interestingly enough a number of the print-outs were 'screenshots' of the organization's internal CRM system, indicating an ad-hoc response.

Data Accuracy and Completeness. In a majority of cases we were able to think of categories of data that the entity holds and did not include in their response. In one instance, the data was fully wrong[2].

Who the Data is Shared With. Only one organization, the city of the Hague, included in their results a list of all entities they had shared our citizen information with. They provided a log with the time and name of the agency requesting the information.

3. A CROWD-SOURCED DATASET OF DATA & PRIVACY PRACTICES

Our experience and those of others (such as [6]) show that, despite some practical hurdles, the right of access can offer insights into data collection and management practices of organizations. This is why we propose a platform to crowd-source and crowd-interpret access requests. Crowd-sourcing allows us to scale to more organizations, check whether collection correlates to a user's background or technical skills, increase validity of the findings, and encourage societal debate.

We have received seed funding and IRB approval to start a pilot project. We plan to recruit approximately 100 Dutch residents as volunteers to send access requests to a number of organizations. Bits of Freedom, a Dutch digital rights NGO, has been offering an online tool named 'Privacy Inzage Machine' [7] to help citizens create access requests, but they did not track whether users submitted the requests. Our platform adds request tracking, and allows participants to share access responses with other participants and researchers on the platform.

We are currently developing the platform specifications. A key issue here is how to take into account privacy considerations. Some important questions include how to allow redacting of shared information, how to guide volunteers to encode received information, and what are the best methods for enabling the final dataset to be shared with external researchers. Table 2 presents data sharing options we envisage. We are considering differential privacy [8] or a mediated framework for sharing the dataset in a privacy-preserving manner with other PETS researchers. We will include "response evaluation", surveying participants on the correctness, completeness, and "creepiness" of received data.

Given that we are currently in the design phase of the platform, the community's suggestions and feedback will be very helpful to us.

Table 2. Data Sharing on the Access Request Platform

	User	PI & admins	Others Users	External Researchers
User info	☑	☑	User choice	☒
Full access response	☑	☒	☒	☒
Redacted access response	☑	☑	User choice	In privacy-preserving way & with extra consent
Process time & info	☑	☑	User choice	*(same)*
Response evaluation	☑	☑	User choice	*(same)*

4. ETHICAL CONSIDERATIONS

Ad-hoc and delayed responses suggest that answering access requests might be a burden for organizations. One could ask whether it is fair for researchers to pose such costs on organizations? While recognizing the dilemma, we believe this is ethically acceptable. First, response costs are mainly due to organizations not having processes in place to handle such requests. One could argue that the requests act as a learning opportunity—and some organizations have indeed indicated this to us in informal follow ups. Second, access requests are an existing legal obligation, and should be seen as a cost of doing business. A cost that European law-makers have deemed necessary to ensure democratic rights.

5. ACKNOWLEDGMENTS

We thank Princeton University's Center for Information Technology Policy (CITP) for providing a seed grant for our pilot, and the NGO Bits of Freedom for advising us on the project.

6. REFERENCES

[1] Mayer-Schönberger, V. and Cukier, K. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think.* Houghton Mifflin Harcourt.

[2] Zuboff, S. 2015. "Big Other: Surveillance Capitalism and the Prospects of an Information Civilization." *Journal of Information Technology*. 30, 75–89. DOI=10.1057/jit.2015.5.

[3] European Union. 1995. Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Official Journal* L 281, 23.11.1995, p31–50.

[4] European Union. 2016. "Regulation 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data". *Official Journal* L 119, 4.5.2016, p. 1–88.

[5] Schrems, M. (2016). "Europe v Facebook." https://europe-v-facebook.org.

[6] Grogan, S. and McDonald, A.M. 2016. "Access Denied! Contrasting Data Access in the United States and Ireland." *Proceedings on Privacy Enhancing Technologies.* 2016 (3). p191–211. doi:10.1515/popets-2016-0023.

[7] Bits of Freedom. nd. *"Privacy Inzage Machine"*. https://www.bof.nl/ons-werk/privacy-inzage-machine/

[8] Dwork, C. 2006, "Differential privacy," in *Automata, Languages and Programming*

[1] See: https://twitter.com/hadi_a/status/840860230896500736

[2] See: https://twitter.com/hadi_a/status/854997285805203457