# An Empirical Analysis of Website Data Deletion and Opt-Out Choices

**Hana Habib**
Carnegie Mellon University
Pittsburgh, PA 15213, USA
htq@cs.cmu.edu

**Yixin Zou**
University of Michigan
Ann Arbor, MI 48109, USA
yixinz@umich.edu

**Aditi Jannu**
Carnegie Mellon University
Moffett Field, CA 94035
aditi.jannu@sv.cmu.edu

**Chelse Swoopes**
Carnegie Mellon University
Pittsburgh, PA 15213, USA
cswoopes@andrew.cmu.edu

**Alessandro Acquisti**
Carnegie Mellon University
Pittsburgh, PA 15213, USA
acquisti@andrew.cmu.edu

**Lorrie Faith Cranor**
Carnegie Mellon University
Pittsburgh, PA 15213, USA
lorrie@cs.cmu.edu

**Norman Sadeh-Koniecpol**
Carnegie Mellon University
Pittsburgh, PA 15213, USA
sadeh@cs.cmu.edu

**Florian Schaub**
University of Michigan
Ann Arbor, MI 48109, USA
fschaub@umich.edu

## Abstract

With the implementation of the GDPR, users are expected to have more choices regarding their online privacy. However, it is important to understand the current paradigm of how these online privacy choices are offered by websites, and evaluate whether people can use them effectively. In this workshop paper we present our progress on an empirical study of websites' data deletion and opt-out choices. We find that although privacy choices are being offered, there is variation across websites in terms of where those choices are located and how they are described. Results from our analysis could help pinpoint these inconsistencies and improve the overall user experience for expressing these types of privacy controls.

## Author Keywords
GDPR, opt-out, privacy, regulation, usability

## ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous; See [http://acm.org/about/class/1998/]: for full list of ACM classifiers. This section is required.

## Introduction
One of the broader goals of the General Data Protection Regulation (GDPR) is to provide Internet users with simpler privacy controls, especially over the data collected about

them from websites [9]. Examples of these privacy controls are higher standards for obtaining consent from users prior to some forms of data collection, and the requirement that users must be able to withdraw their consent, or opt out, at any time. Another GDPR requirement grants Internet users the "right to be forgotten," or the ability to request websites to delete their personal data.

The new GDPR requirements are an effort to address data collection practices that many Internet users find concerning. One such practice is targeted advertising, which allows websites to serve users advertisements tailored to their previous online activities [8, 11]. While there are tools and mechanisms that allow users to opt out of receiving this type of advertising, people find them difficult to understand and use [7, 8]. Beyond choices regarding targeted advertising, websites also often provide opt-out mechanisms for receiving marketing communications, as required by law in the United States [4].

In order for these privacy controls to be effective, websites and device manufacturers must ensure they are usable by their consumers. Privacy choice mechanisms must also be adaptable to a variety of contexts, from websites with user profile information to wearable or Internet of Things (IoT) devices. While some of these tools are improving, it is still difficult for users to express privacy choices on websites. Furthermore, there is little standardization of how or where these controls are presented to users.

We present our progress on an analysis of Internet privacy choices in which we collect empirical metrics about data deletion and opt-out mechanisms currently offered on websites. The purpose of this analysis is to better understand current practices and to ultimately design better consent and opt-out experiences, which are explicitly related to GDPR requirements. This census study will be followed by future work that will investigate usability issues related to these existing mechanisms. Lastly, we will incorporate our findings to the design and evaluation of an adaptable choice and consent platform that allows websites to offer more usable privacy choices.

## Background & Related Work

In this section we provide some background and present prior work related to the privacy choices that are the focus of this analysis: data deletion options and email and targeted advertising opt-outs. We also provide an overview of related legislation and industry self-regulatory requirements.

With respects to data deletion, the GDPR grants European Union citizens the "right to be forgotten," which stipulates that, under certain conditions, companies must comply with requests from users to erase their personal data [9]. One such implementation of the "right to be forgotten" would be through account deletion requests, or the ability for users to delete certain information related to their profile.

Pertaining to email communications, the United States' Controlling the Assault of Non-Solicited Pornography And Marketing (CAN-SPAM) Act of 2003 requires businesses that send commercial email messages a means for recipients to opt out of receiving them [4]. As a result, websites have started offering consumers increased controls for the email messages they receive. A recent audit by the Online Trust Alliance found that 92% of the retail websites surveyed offered "unsubscribe" links within the message [3]. Some websites, such as Amazon, also offer email options through the user account settings.

Related to data collection for targeted advertising, the GDPR requires websites to obtain active agreement from users and allow users to withdraw their consent at any time. This

Step 3: Look for an "about advertising" or "ad choices" related link on the home page. Click on the "about advertising" or "ad choices" link if it is there.

Is there an "about advertising" or "ad choices" related link on the home page?

| Yes, and it works |
|---|
| Yes, but it's broken |
| No |

**Figure 1:** One question related to targeted advertising opt-outs included in our data collection template.

active agreement comprises of an affirmative action and does not include pre-checked boxes or idleness from the users. Moreover, specific consent for each data processing operation must be provided in an easily acceptable form using simple, clear language and visualization if need be. If the user is a child, the language must be understandable by a child. The GDPR also restricts processing of some categories of personal data, despite having obtained explicit consent [9].

Prior consent requirements impacting the advertising industry included guidelines set by the Digital Advertising Alliance (DAA) and Network Advertising Initiative (NAI), two industry self-regulatory groups formed by online advertisers [2, 5]. However, users face a variety of challenges in opting out of targeted advertising. A 2011 study by Komanduri et al. found infrequent compliance to the NAI enhanced choice requirements [1]. Furthermore, other research has shown some advertising opt-out tools to be more effective than others [6], and has revealed usability issues related to several of these tools [7, 8].

Our analysis aims to construct a census of all three of these privacy choices. We will both examine the state of compliance to current and upcoming regulation, and develop a taxonomy of how these privacy choices are currently being offered.

## Methodology

Our methodology consists of a manual analysis of data deletion and opt-out choices offered on websites, as well as an automated analysis of privacy policies.

For the manual analysis, we have developed an annotation template to collect metrics about the data deletion, email, and targeted advertising choices offered by these websites. For all choices, we record information such as where the

privacy choice is located, the shortest path to it as measured by numbers of clicks, and other information about it described in privacy policy. Figure 1 shows one of the questions included in our data analysis template related to targeted advertising opt-outs.

To make sure our analysis is comprehensive and unbiased, we developed the template in an iterative manner, and follow a protocol that standardizes the procedures for analyzing each website. A group of four researchers worked on the initial versions of the template, each coding a small set of websites independently, then reconciling the results, and revising the template collaboratively.

For a larger scale, automated analysis of privacy policies, we will use machine learning based tools developed by researchers as part of the Usable Privacy Policy Project.[1] Specifically, we will leverage prior work by Sathyendra et al. [10], and the project's Explore website. [2] With these policy analysis tools we aim to verify and scale certain metrics from our manual analysis, such as whether opt-out or data deletion choices exist in the privacy policy, where in the policy they are described, and what additional information about them is provided to users. Using these tools will allow us to better understand how data deletion and opt-out choices are currently presented in privacy policies, across a broader spectrum of websites.
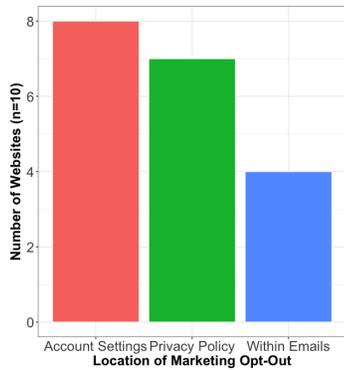
## Preliminary Results

In this section we present some insights from piloting our data collection template with ten of the most popular websites in the United States, based on Amazon Alexa's ranking of top sites.[3] The ten websites included in our pilot were
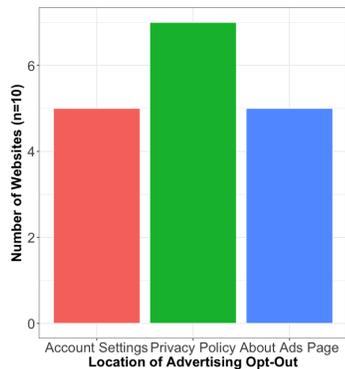
---

[1] Usable Privacy Policy Project: https://www.usableprivacy.org
[2] Usable Privacy Explore: https://explore.usableprivacy.org/
[3] Amazon Alexa Top Sites: https://www.alexa.com/topsites

**Figure 2:** Eight of the ten websites provided an opt-out for marketing or email communications. Four websites offered opt-outs in all three locations.



**Figure 3:** Seven of the ten websites provided an opt-out for targeted advertising. Four websites offered advertising opt-outs in all three locations.

Amazon, Craigslist, eBay, ESPN, Facebook, Google, Tumblr, Twitter, Wikipedia, and Yahoo. These preliminary results are from only a subset of the data collected through our analysis template. Additionally, since this pilot sample consisted of only popular websites, these results are likely not representative of the privacy choices available on less trafficked websites, and the Internet as a whole. Overall, we find that privacy choices are available on the majority of these websites, but websites vary in how they present them to users.

*Email Related Opt-Outs*
Eight of the ten websites evaluated offered opt-outs for marketing or email communications in the account settings. Most of the opt-outs offered were related to marketing from the website, but some websites also included options to opt-out of newsletters and activity updates. One website in our analysis, Craigslist, does not send users any marketing communications. Of these eight websites that offered email related opt-outs, seven included information about opting out of communications in their privacy policy. Four websites also specified that email communications will contain an "unsubscribe" function within the emails. Figure 2 provides a summary of where these opt-outs are located on the website.

Looking at where opt-outs for email communications are described in the privacy policy, four of the seven websites containing email opt-outs in the privacy policy described these opt-outs under a section in the policy where the title indicated "choice" or "preference," while two explicitly used the word "email" in section title. Different from the others, Twitter described email opt-outs in their privacy policy under a section titled "Information Collection and Use."

*Advertising Related Opt-Outs*
We analyzed the same ten websites for their opt-out offerings related to targeted advertising. Seven websites presented information about advertising opt-outs in their privacy policy. Five of the ten websites presented advertising opt-outs through an informational page about advertising practices (e.g., "About Ads") linked to the home page, and five included advertising opt-out settings in the user account settings. Four websites offered advertising opt-outs through all three means. Wikipedia and Craigslist indicated in their privacy policies that they do not serve targeted ads. Figure 3 provides a summary of where advertising opt-outs are located on the website.

Of the seven websites that offered advertising related opt-outs, six had a site-implemented adverting opt-out function. Three of these six also offered opt-outs through either the DAA or NAI opt-out tools, while one website (ESPN) only provided links to the DAA opt-out service.

Within the privacy policy, there was variation in where advertising opt-outs were described. Four of the websites described them in sections with titles related to "choice," two included "information" in the section title, while one website (Yahoo) described them under "Personalized Experience."

*Data Deletion Options*
In total, seven of the ten websites analyzed provided at least one mechanism for users to delete their account or information related to their account. Four websites offered account deletion mechanisms in both the privacy policy and user account settings, two provided them only in the account settings, while Yahoo provided account deletion options only in the privacy policy. Figure 4 provides a summary of where these opt-outs are located on the website.
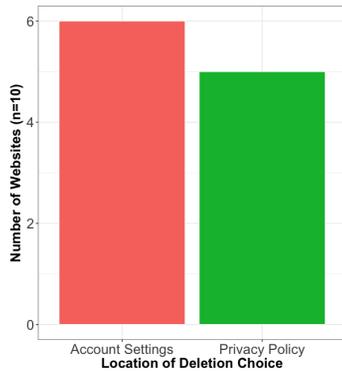
**Figure 4:** Seven of the ten provided at least one way for users to delete their account or account data. Four websites offered opt-outs through both the privacy policy and account settings.

Four of the five websites described account deletion mechanisms in their privacy policy under sections that included the words "delete" or "modify," while ESPN's policy described it under "Your Controls and Choices."

## Discussion & Future Work

The results from our preliminary analysis suggest that privacy choices related to data deletion and consent are present on popular websites, but may not be all that usable. For example, some websites may only offer advertising opt-outs through their privacy policy, while others may only do so through the user account settings. Even within the privacy policy, the terms used to describe sections containing opt-outs seem inconsistent from website to website. These types of inconsistencies likely make it harder for users to find privacy choices and act on them. We aim to develop a standardized platform for privacy choices which will streamline this process for users.

Our next step will be to analyze the offerings of privacy choices of more websites using the annotation template and automated analysis tools. We will primarily focus on Alexa global top 500 sites, since these websites impact the majority of Internet users. We also plan to include a smaller sample, randomly selected from the Alexa top 500 through 10,000 websites to explore whether the privacy choices vary across websites with different volumes of traffic.

Following this analysis, we would like to investigate the usability aspects of these opt-outs. One direction would be to examine whether these opt-outs are functioning in the way as claimed by the website or third-parties from a technical perspective. Another direction would be to explore whether these opt-outs are conceptualized and utilized differently by individual end-users with different backgrounds and characteristics, such as those with cybersecurity knowledge or higher level of privacy concern.

Finally, based on our findings from the empirical analysis and usability studies, we will prototype alternatives to current consent and control offerings, and evaluate their effectiveness. Our ultimate goal is to develop a more usable opt-out platform for websites to adopt.

## Acknowledgements

## REFERENCES

1. Adchoices-Compliance with Online Behavioral Advertising Notice and Choice Requirements, author=Komanduri, Saranga and Shay, Richard and Norcie, Greg and Ur, Blase, journal=ISJLP, volume=7, pages=603, year=2011, publisher=HeinOnline. (????).

2. 2009. Self-Regulatory Principles for Online Behavioral Advertising. (Jul 2009). `http://digitaladvertisingalliance.org/principles`

3. 2017. 2017 Email Marketing & Unsubscribe Audit. (Dec 2017). `https://www.ftc.gov/tips-advice/business-center/guidance/can-spam-act-compliance-guide-business`

4. 2017. CAN-SPAM Act: A Compliance Guide for Business. (Mar 2017). `https://otalliance.org/system/files/files/initiative/documents/2017emailunsubscribeaudit.pdf`

5. 2018. 2018 NAI Code of Conduct. (2018). `https://www.networkadvertising.org/sites/default/files/nai_code2018.pdf`

6. Rebecca Balebako, Pedro Leon, Richard Shay, Blase Ur, Yang Wang, and L Cranor. 2012. Measuring the Effectiveness of Privacy Tools for Limiting Behavioral Advertising. Web.

7. Pedro Leon, Blase Ur, Richard Shay, Yang Wang, Rebecca Balebako, and Lorrie Cranor. 2012. Why Johnny can't opt out: a usability evaluation of tools to limit online behavioral advertising. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 589–598.

8. Aleecia M McDonald and Lorrie Faith Cranor. 2010. Americans' attitudes about internet behavioral advertising practices. In *Proceedings of the 9th annual ACM workshop on Privacy in the electronic society*. ACM, 63–72.

9. EU GDPR Portal. 2017. Home Page of The European Union General Data Protection Regulation (GDPR). (2017). `https://www.eugdpr.org/eugdpr.org.html`.

10. Kanthashree Mysore Sathyendra, Shomir Wilson, Florian Schaub, Sebastian Zimmeck, and Norman Sadeh. 2017. Identifying the Provision of Choices in Privacy Policy Text. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 2774–2779.

11. Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. 2012. Smart, useful, scary, creepy: perceptions of online behavioral advertising. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*. ACM, 4.