Designing Visualization and Exploration Tools for Data Access Under GDPR/CCPA

Sophie Veys, Madison Stamos, Nathan Reitinger[†], Michelle L. Mazurek[†], and Blase Ur
University of Chicago and [†]University of Maryland
Email: {sophiev,maddiestamos,blase}@uchicago.edu, {nlr,mmazurek}@cs.umd.edu

Abstract—In accordance with GDPR and CCPA data-access rights, many companies now allow users to download or view their data related to a particular website or app. However, these data files are often loaded with jargon, poorly organized, and altogether difficult for users to understand. To help users better understand their data, we hope to generate ideas for data visualization tools via site-specific participatory-design sessions. We hope that our findings will inform the design of such tools, which could make data collection and use more transparent.

I. Introduction

The EU General Data Protection Regulation (GDPR) [1] and the California Consumer Privacy Act (CCPA) [2] are two of the most prominent data-protection laws. Both give consumers legal rights in many circumstances to access personal data collected or processed by a regulated entity.[1] We term the data provided in response a *data download*. Fulfilling such a request, though, does not require providing more than *raw* data. In fact, data-portability requirements[2] make it likely for entities to argue that raw data is the most appropriate response.

Companies have little incentive to make data downloads user-friendly. Some might even deliberately obscure these files to hide unsavory practices. As a result, it falls upon consumers to decipher their own data downloads. Doing so is challenging due to the use of technical language, lack of explanations, and general disorganization. For instance, many files are in JSON or plain text formats. The timestamps in Facebook data JSON files are in UNIX, not human-recognizable, format.

It is not enough to simply release data downloads if consumers cannot easily understand what they mean. At a high level, we argue that policymakers should clarify readability requirements. Regardless, we feel that the availability of this hard-to-decipher data presents an opportunity for privacy technologists to develop tools that equip consumers to better understand their data downloads. **Through participatory design sessions based in data downloads from exercising GDPR/CCPA access rights, we will study consumers' understanding of data download formats and shape the design of new tools for visualizing data downloads.**

II. Approach

We will identify popular, non-embarrassing websites that offer data downloads, clustering by category (e.g., social media) and choosing one representative site from each. We will recruit participants in groups of 4–6 for hour-long participatory design sessions. So that designs can be site-specific, each session will focus on one site. Participants will be asked to download their data ahead of time and bring it to the session. Sessions will include the following six activities, designed to encourage discussion and inspire ideas about ideal visualizations:

Data Visualization 101. To inform and inspire participants, we will show examples of data-visualization techniques.

Scavenger Hunt. To familiarize participants with their data downloads, we will provide a list of items to find in their data, such as a deleted message or the timestamp of a purchase.

Questions. To gauge participants' understanding of their data downloads, participants will answer questions from the moderator about the content and format of their data files, as well as general questions about data-collection practices.

Highlighting. Participants will highlight the categories of information from the data downloads (e.g., login locations) they would be most interested in seeing visualized.

Post-It Ideation. We will give participants post-its and ask them to write down all design features or data categories that would be included in their ideal visualization tool. As a group, they will select the most important post-its, allowing participants to share ideas and refine their designs.

Sketch Activity. Participants will individually sketch their ideal version of a data visualization tool, which could include content, format, and menu options.

We will analyze and aggregate the data generated from each activity to identify participants' preferred content and format for data visualizations. Based on this list, we will propose a blueprint for what a data visualization tool might look like. While each session is site-specific, we hope that many of the features may generalize across sites and categories. We subsequently intend to create our own data visualization tool exhibiting these characteristics. This tool will accept data downloads from many different websites, visualizing and cross-referencing a consumer's collection of data downloads.

III. Related Work

Many researchers have created data visualization tools to help make information more digestible [3], [4]. Almuhimedi et al. found that users took steps to protect their privacy when made aware of how their data was being used [5]. Using the participatory design methodology laid out by Spinuzzi [6], we hope to create a tool, designed in part by consumers themselves, that provides greater data transparency.

---

[1]GDPR Article 15 [1]; CCPA §§1798.100, 1798.130, 1798.140 [2].
[2]GDPR Article 20 [1]; CCPA §1798.100(d) [2].

REFERENCES

[1] "General Data Protection Regulation," April 2016, (accessed January 2020).

[2] "California Consumer Privacy Act of 2018," 2018, (accessed January 2020).

[3] J. Ahrens, B. Geveci, and C. Law, "Paraview: An end-user tool for large data visualization," *Technical Report*, 2005.

[4] H. Jiang, F. Wang, N. P. Dyer, and W. H. Wong, "Cisgenome browser: a flexible tool for genomic data visualization," *Bioinformatics*, vol. 26, no. 14, pp. 1781–1782, July 2010.

[5] H. Almuhimedi, F. Schaub, N. Sadeh, I. Adjerid, A. Acquisti, J. Gluck, and L. F. Cranor, "Your location has been shared 5,398 times!: A field study on mobile app privacy nudging," in *Proc. CHI*, April 2015, pp. 787–796.

[6] C. Spinuzzi, "The methodology of participatory design," *Technical Communication (Washington)*, vol. 52, no. 2, pp. 163–174, May 2005.