

Pursuing Usable and Useful Data Downloads Under GDPR/CCPA Access Rights via Co-Design

*Sophie Veys, Daniel Serrano, Madison Stamos, Margot Herman,
Nathan Reitinger, Michelle L. Mazurek, and Blase Ur*



THE UNIVERSITY OF
CHICAGO



UNIVERSITY OF
MARYLAND





Right of Access



Data Downloads



```
{
  "endTime" : "2019-09-16 05:16",
  "artistName" : "Chris Cornell",
  "trackName" : "Call Me A Dog - Live At Queen",
  "msPlayed" : 19349
},
{
  "endTime" : "2019-09-16 05:30",
  "artistName" : "Guns N' Roses",
  "trackName" : "Knockin' On Heaven's Door",
  "msPlayed" : 67968
},
{
  "endTime" : "2019-09-16 05:33",
  "artistName" : "Skid Row",
  "trackName" : "18 and Life",
  "msPlayed" : 79850
},
{
  "endTime" : "2019-09-16 05:38",
  "artistName" : "Third Eye Blind",
  "trackName" : "Semi-Charmed Life",
  "msPlayed" : 1728
},
},
```

My Own Spotify Data

AirDrop

Recents

Desktop

Applications

Documents

Downloads

iCloud

iCloud Drive

Locations

Network

Tags

vegetable

test

Blue

facebook-sophieveys188

Search

Name	Date Modified	Size	Kind
apps_and_websites.html	Yesterday at 11:03 AM	65 KB	HTML document
> your_off-facebook_activity	Yesterday at 11:03 AM	--	Folder
your_off-facebook_activity.html	Yesterday at 11:03 AM	59 KB	HTML document
> bug_bounty	Yesterday at 11:03 AM	--	Folder
no-data.txt	Yesterday at 11:03 AM	32 bytes	Plain Text
> campus	Yesterday at 11:03 AM	--	Folder
campus_email_info.html	Yesterday at 11:03 AM	25 KB	HTML document
> comments_and_reactions	Yesterday at 11:03 AM	--	Folder
comments.html	Yesterday at 11:03 AM	110 KB	HTML document
posts_and_comments.html	Yesterday at 11:03 AM	484 KB	HTML document
> events	Yesterday at 11:03 AM	--	Folder
event_invitations.html	Yesterday at 11:03 AM	40 KB	HTML document
your_event_responses.html	Yesterday at 11:03 AM	29 KB	HTML document
> facebook_accounts_center	Yesterday at 11:03 AM	--	Folder
accounts_center.html	Yesterday at 11:03 AM	26 KB	HTML document
> facebook_gaming	Yesterday at 11:03 AM	--	Folder
> facebook_marketplace	Yesterday at 11:03 AM	--	Folder
> facebook_payments	Yesterday at 11:03 AM	--	Folder
> friends_and_followers	Yesterday at 11:03 AM	--	Folder
> fundraisers	Yesterday at 11:03 AM	--	Folder
> groups	Yesterday at 11:03 AM	--	Folder
index.html	Yesterday at 11:03 AM	78 KB	HTML document
> journalist_registration	Yesterday at 11:03 AM	--	Folder
> likes_and_reactions	Yesterday at 11:03 AM	--	Folder
> location	Yesterday at 11:03 AM	--	Folder
> messages	Today at 9:36 AM	--	Folder
> music_recommendations	Yesterday at 11:03 AM	--	Folder
> news	Yesterday at 11:03 AM	--	Folder
> notifications	Yesterday at 11:03 AM	--	Folder
> other_activity	Yesterday at 11:03 AM	--	Folder
> other_logged_information	Yesterday at 11:03 AM	--	Folder
> other_personal_information	Yesterday at 11:03 AM	--	Folder
> pages	Yesterday at 11:03 AM	--	Folder
> people_and_friends	Yesterday at 11:03 AM	--	Folder
> photos_and_videos	Yesterday at 11:02 AM	--	Folder
> polls	Yesterday at 11:03 AM	--	Folder
> posts	Yesterday at 11:03 AM	--	Folder
> preferences	Yesterday at 11:03 AM	--	Folder
> privacy_checkup	Yesterday at 11:03 AM	--	Folder
> profile_information	Yesterday at 11:03 AM	--	Folder
> reviews	Yesterday at 11:03 AM	--	Folder
> saved_items_and_collections	Yesterday at 11:03 AM	--	Folder
> search	Yesterday at 11:03 AM	--	Folder

Research Questions

RQ1

Reactions to format and
content of data downloads

RQ2

What information is important?
What uses are imagined?

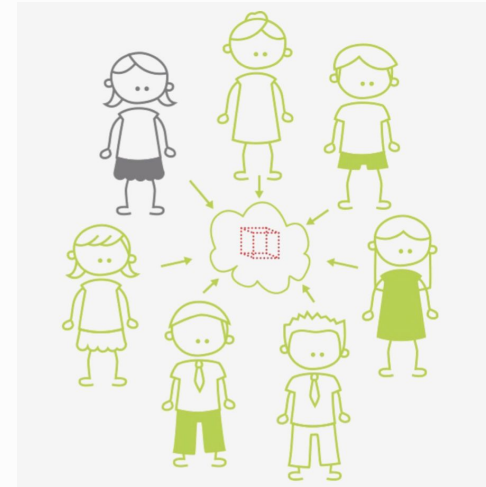
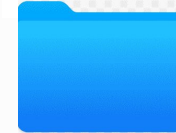
RQ3

How could data downloads
be redesigned?

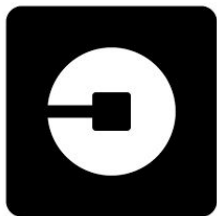
Study Design

12 Virtual Focus Groups

42 Participants



Company Selection




Activities

GDPR/CCPA

101



File Exploration

 **Your Off-Facebook Activity**
Your activity from the businesses and organizations you visit off of Facebook
[View on Facebook](#)

blushmark.com

wish.com

shein.com

intuit.com


Trivia Crack

2048

Starbucks

Grubhub: Local Food Delivery

Venmo



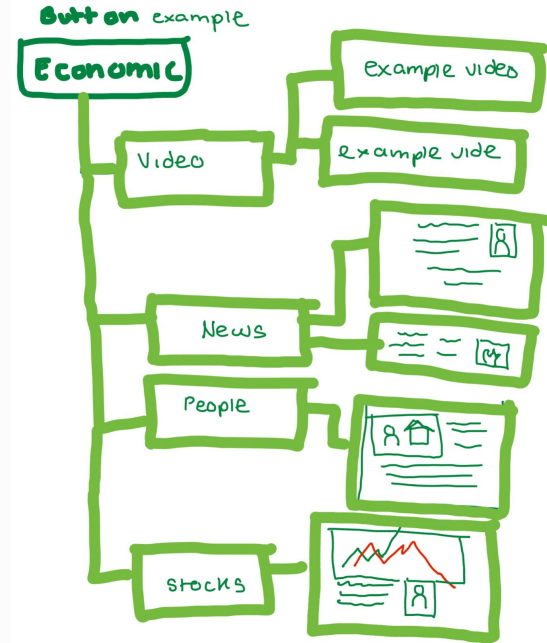
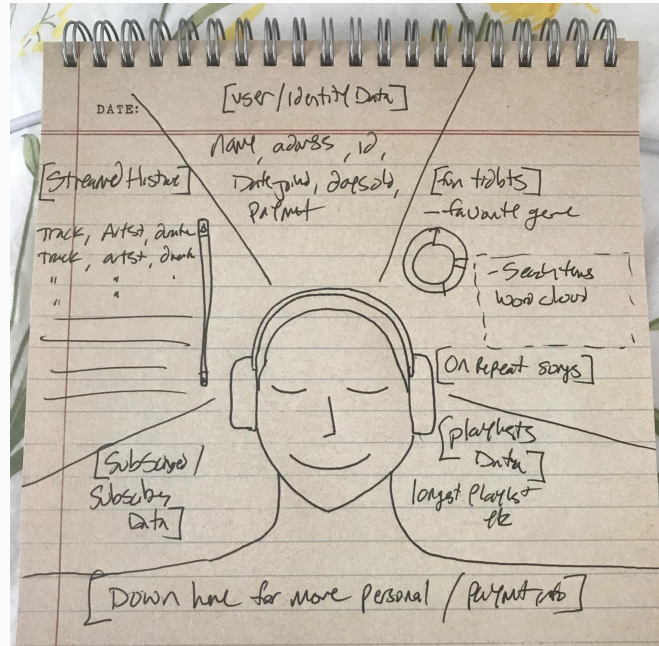
Discussion



Data Viz 101



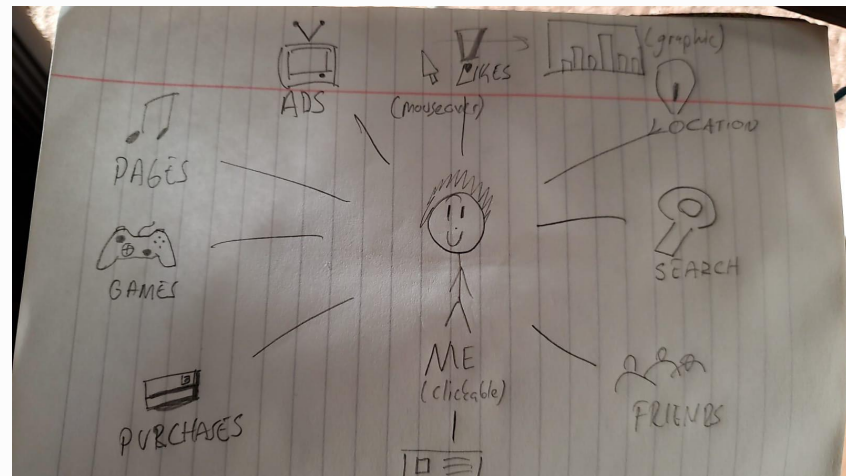
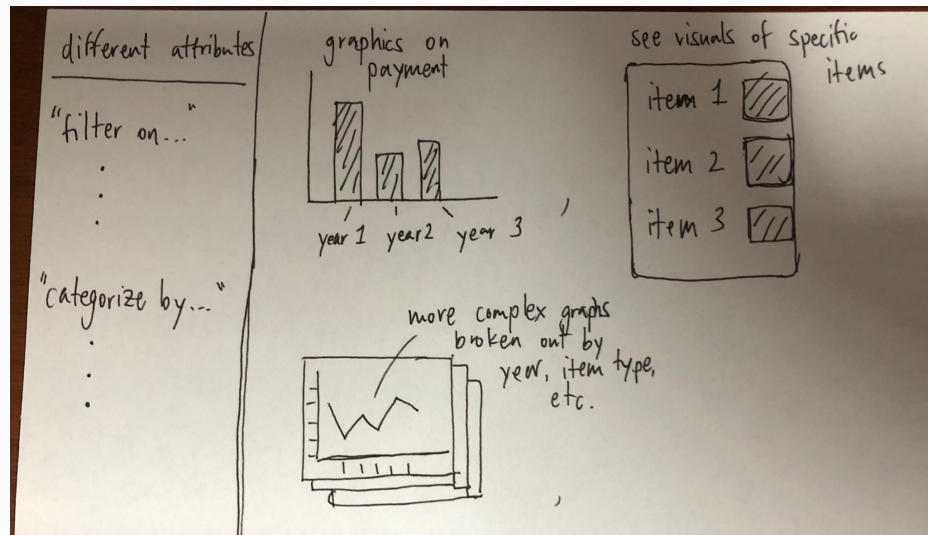
Sketch Activity





Data Analysis

Sketch Analysis



Pursuing Usable and Useful Data Downloads Under GDPR/CCPA Access Rights via Co-Design

Sophie Veys, Daniel Serrano, Madison Stamos, Margot Herman,
Nathan Reitering¹, Michelle L. Mazurek², Blase Ur
University of Chicago, ¹University of Maryland

Abstract

Data privacy regulations like GDPR and CCPA define a *right of access* empowering consumers to view the data companies store about them. Companies satisfy these requirements in part via *data downloads*, or downloadable archives containing this information. Data downloads vary in format, organization, comprehensiveness, and content. It is unknown, however, whether current data downloads actually achieve the transparency goals embodied by the right of access. In this paper, we report on the first exploration of the design of data downloads. Through 12 focus groups involving 42 participants, we gathered reactions to six companies' data downloads. Using co-design techniques, we solicited ideas for future data download designs, formats, and tools. Most participants indicated that current offerings need improvement to be useful, emphasizing the need for better filtration, visualization, and summarization to help them hone in on key information.

1 Introduction

The principle of **data access** states that subjects should be able to obtain a copy of the data that has been collected about them. For decades, this principle has appeared in information privacy frameworks [24]. For example, access is one of the five core facets of the U.S. Federal Trade Commission's Fair Information Practice Principles (FIPPs) [24]. In past decades, while other FIPPs directly impacted consumers (e.g., the principle of notice underpins the ubiquity of privacy policies [66]), the principle of access was mostly ignored. In recent years, however, rights of access have been strengthened. In the Eu-

ropean Union, Article 15 [79] of the General Data Protection Regulation (GDPR) enshrines a "right of access by the data subject." Similarly, under the California Consumer Privacy Act (CCPA), businesses must respond to consumer "requests to know" about data collected about them, enabling them "to access, view, and receive" a copy of that data [76].

Consumers might want access to their data for many reasons. First, data downloads can help users uncover distressing aspects of the online data ecosystem. Prior work has found that consumers can feel uneasy upon seeing evidence of online tracking and data collection [78, 81, 88]. Further, consumers often become upset when they feel that data has been misused or taken out of context [52], including for advertising [27] or politics [34]. In a widely-discussed article, Hill used data downloads to expose "secret consumer scores" in which consumers' purchase histories and demographics impact their eligibility for refunds [31]. Access to data is a prerequisite for consumers to modify any incorrect information (the privacy principle of participation) [24]. Additionally, awareness of data collection might encourage users to exercise their right of erasure [9] or motivate other privacy-protective actions.

Privacy concerns aside, there are more practical reasons consumers might want access to their data. Many consumers have data spread across many platforms. For example, a consumer might have pictures published to Twitter, Instagram, and Tumblr. In the event they lose the device on which the original pictures are stored, they might try to reclaim as many photos as possible. Alternatively, a consumer might wish to move from one service (e.g., Spotify) to a competitor (e.g., Amazon Music), yet wish to seamlessly transfer their carefully curated playlists and other personal data. The pursuant right of **data portability**, which enables consumers to transfer personal data across services via interoperable formats, is also enshrined in both GDPR [79] and CCPA [76].

To comply with these legal rights of data access and portability, many companies have begun to offer what we term **data downloads**, which are either files or archives of files containing the identifiable data a business or other data processor has collected about a consumer. Figure 1 shows ex-

Copyright is held by the author/owner. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee.
USENIX Symposium on Usable Privacy and Security (SOUPS) 2021, August 8–10, 2021, Virtual Conference.

Key Findings

**“It’s like they
didn’t even try.
They just kind
of dumped it on
you.”**



**“Most of the interesting
data is stored in these
files, that as a
non-specialist, I can’t
read... We’re effectively
illiterate when it comes
to reading this
additional data they’ve
been collecting.”**



Design Suggestions

***Meaningful
Organization**

***Interactivity &
Exploration**

***Filtration**

***Direct
Manipulation**

***Aggregation &
Inferencing**

“What’s interesting to me is how my online behavior is affecting how this company and all the affiliates see me. And in what category, say, they put me or don’t put me... That has a way broader implication... Who is programming these algorithms?... Do they represent a broader part of society or are they all from a very similar group? ”



Design Suggestions

***Meaningful
Organization**

***Interactivity &
Exploration**

***Filtration**

***Direct
Manipulation**

***Aggregation &
Inferencing**

Policy Suggestions

- *Data Access vs. Data Portability

- *Required Content

- *Explanations for Missing Data

Pursuing Usable and Useful Data Downloads Under GDPR/CCPA Access Rights via Co-Design

*Sophie Veys, Daniel Serrano, Madison Stamos, Margot Herman,
Nathan Reitinger, Michelle L. Mazurek, and Blase Ur*



THE UNIVERSITY OF
CHICAGO



UNIVERSITY OF
MARYLAND

