Hidden in Plain Bytes: Investigating Interpersonal Account Compromise with Data Exports

Julia Nonnenkamp
nonnenkamp@wisc.edu
University of Wisconsin–Madison
Madison, Wisconsin, USA

Naman Gupta
n@cs.wisc.edu
University of Wisconsin–Madison
Madison, Wisconsin, USA

Abhimanyu Dev Gupta
adgupta2@wisc.edu
University of Wisconsin–Madison
Madison, Wisconsin, USA

Rahul Chatterjee
chatterjee@cs.wisc.edu
University of Wisconsin–Madison
Madison, Wisconsin, USA

Abstract

When survivors of technology-facilitated abuse (TFA) suspect someone has accessed their online accounts, they often rely on built-in account security interfaces (ASIs), such as trusted device lists within settings, to assess account compromise. However, these interfaces typically offer limited or ambiguous details about past account accesses and security-critical events. Under right of access provisions in data protection laws, users can request structured exports of their personal data from online services. In this study, we explore whether and how data exports can supplement ASIs to support compromise investigations, particularly in interpersonal threat contexts. We simulated four types of account compromise attacks across six popular platforms, analyzing the resulting data exports and ASIs. Our findings show that data exports consistently contain more granular login histories and richer device/network identifiers than interfaces. Some even link security-related actions (e.g., password changes) and other post-authentication activity to specific devices, offering forensic value for identifying compromise. We discuss usability and other practical challenges of using data exports during TFA interventions.

CCS Concepts

• **Security and privacy** → **Human and societal aspects of security and privacy**; • **Applied computing** → *Investigation techniques.*

Keywords

technology-facilitated abuse, intimate partner violence, GDPR, data rights, account takeover, usable security

ACM Reference Format:
Julia Nonnenkamp, Naman Gupta, Abhimanyu Dev Gupta, and Rahul Chatterjee. 2025. Hidden in Plain Bytes: Investigating Interpersonal Account Compromise with Data Exports. In *Proceedings of the 2025 ACM SIGSAC Conference on Computer and Communications Security (CCS '25), October 13–17, 2025, Taipei, Taiwan.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3719027.3765147

CCS '25, Taipei, Taiwan

ACM ISBN 979-8-4007-1525-9/2025/10
https://doi.org/10.1145/3719027.3765147

1 Introduction

As more aspects of our personal and professional lives shift online, our physical and digital safety has become increasingly dependent on the security of our online accounts. This risk is particularly pronounced for individuals experiencing technology-facilitated abuse (TFA) — a spectrum of harmful behaviors such as harassment, surveillance, and coercive control through digital technologies, often carried out by someone known to the survivor (victim) [33, 59, 75]. Most abusers (adversaries) in TFA contexts use technologically simple yet effective methods to gain authenticated access to survivors' accounts, weaponizing their knowledge of survivors' private information (e.g., SSNs, passwords) and physical access to their devices [31–33, 38, 48, 72, 79]. After initial authentication, abusers may exploit unauthorized, privileged account access by modifying critical settings, enabling location sharing, and intercepting messages.

When a survivor suspects their accounts might be compromised, they may visit *account security interfaces* (ASIs) [21] provided by online services on their mobile apps and websites. These interfaces inform users about the state of their account's security by centralizing key *security-related information* (SRI), including active sessions, recent logins, and password changes. Although ASIs are important sources of information for survivors, Daffalla et al. [21] identified serious limitations with respect to the reliability and completeness of the information displayed within them. Many ASIs provide only coarse-grained device identifiers (e.g., "iPhone"), which can make it difficult for survivors to differentiate between suspicious and benign devices. Others show only currently active sessions instead of historical logins, and security logs may be time-limited [21].

In this study, we explore an alternative approach to investigating account compromise: *utilizing data exports obtained from online services.* Data protection laws such as the GDPR in the European Union [27], CCPA in California, USA [11, 12], and DPDPA in India [40] grant individuals the right to obtain copies of their personal data from companies and organizations [40, 42, 60]. Most large, online services have developed tools for users to request and download archives of their own data, which we will refer to as *data exports.* While laws and platforms vary in what they consider to be "personal" data, existing research [61, 70] and our preliminary analysis suggest that data exports contain useful SRI for identifying and explaining instances of account compromise. However, the syntactic and semantic structures of data exports are often not well-documented by the services that provide them [9, 82], making it hard to extract all relevant

user information. Moreover, no prior work has studied whether the SRI in data exports is more comprehensive than what ASIs provide. Therefore, we consider the following three research questions:

- ***RQ1*** *What types of information present in data exports are relevant to investigating account compromise by an interpersonal abuser?*
- ***RQ2*** *How does security-related information (SRI) within data exports compare with information presented in account security interfaces?*
- ***RQ3*** *To what extent can we map data exports to abuser actions during account compromise attacks?*

We examined data exports from six popular online services: two services provided by operating system vendors (Apple/iCloud and Google) and four social media & messaging services (Facebook, Instagram, Snapchat, and Discord). Because data exports contain such sensitive information [6, 35, 64], we created researcher-controlled accounts and simulated both benign and adversarial usage for over three months (see Figure 1). We simulated four types of attacks requiring authenticated access to the survivor's accounts: (1) account surveillance, (2) location monitoring, (3) impersonation, and (4) lockout & control. We downloaded the data exports twice during the study period and designed a unified data model and parser for files from all six services.

In Section 4, we developed a label system for data exports that characterizes both security-related information (authentication and security-critical settings) and additional user activity (e.g., messages, interactions). In Section 5, we conducted focused walkthroughs of ASIs from our six services to compare their provided SRI with the SRI in data exports. Finally, in Section 6, we mapped simulated attack steps to specific files and components within data exports.

Thorough account compromise investigations are essential for TFA survivors — not only to understand how an attack occurred, but also to plan for their future safety. Our study demonstrates that data exports may be a valuable resource in such investigations. We urge the research community and industry partners to explore ways to operationalize data export analysis by developing tools that support automated detection and reconstruction of attacks. We provide supplementary materials (parser code & documentation) on GitHub.[1] Our dataset of data exports is available on request via Zenodo.[2]

Contribution. The main contributions are the following:

- *Characterization of security-related information in data exports.* We locate SRI, including authentication and security-critical settings changes, that can be used to identify potential account compromise. Additionally, we highlight other valuable activity data for this purpose, such as messages, searches, and click analytics.
- *Comparison with ASIs.* We show that data exports provide strictly more SRI than ASIs for all services. Exports provide more fine-grained information about device identifiers; more complete records of past logins and security-critical settings changes; and longer log histories compared to ASIs.
- *Identification of specific abuser actions.* We successfully reconstruct login sequences from our attack simulations using data exports from all platforms studied. In addition, we demonstrate our ability to use session-level metadata from Discord and Google to map post-authentication activity on those apps, which is impossible using ASIs alone.

[1] https://github.com/WISPR-lab/data-exports-tfa. We also provide parser utilities as supplemental materials in a full version of this paper available on the 1st author's website.
[2] https://doi.org/10.5281/zenodo.17058860

2 Background

In this section, we first survey work on technology-facilitated abuse mechanisms and interventions, as well as user experiences with security interfaces. Then, we review the legal origins of data exports, their general features, and recent works that use them as a research medium.

2.1 Technology-Facilitated Abuse

Technology-facilitated abuse (often referred to as online abuse [77]) describes deliberate misuse of technology to "stalk, coerce, intimidate, threaten or otherwise harm" another individual [8]. A 2024 study among U.S. college students found that as many as 70% had experienced some forms of TFA, including stalking, online threats, non-consensual intimate imagery, and account takeover [36].

TFA has been studied widely within the context of intimate partner violence (IPV) [8, 16, 33, 37, 72, 86]. Researchers have investigated abuse mechanisms through interviews [8, 32, 33], crawling web content to identify spyware and spy devices [4, 5, 14, 16, 75, 79], and evaluating "anti-security" and surveillance tactics shared on TikTok and Reddit [79, 83]. Thomas et al. [77] developed a taxonomy of TFA-related attacks based on criteria of audience, medium, and abuser capability. Our study focuses on three categories of attacks identified by Thomas et al. [77] that involve different degrees of privileged access to the survivor's online accounts: *impersonation* ("hijacked communication"), *surveillance*, and *lockout & control*.

Interventions for TFA have centered around the *clinical computer security* model introduced by Havron et al. [38], which outlines a framework for providing structured, individualized, and trauma-informed help to survivors of TFA within the IPV context [31, 80]. Three such tech clinics exist in the U.S. [15, 20, 76], providing survivors, or clients, with one-on-one technical support from trained technology consultants. Consultants help clients identify possible spyware and account compromise, proactively enable security features, and assist in safety planning alongside their case manager or advocate [38]. Much of the consultant's work involves using structured protocols [78] to inspect account security settings and helping survivors interpret the presented information.

2.2 Account Security Interfaces

The term *account security interface* (ASI), introduced by Daffalla et al. [21], refers to application user interfaces that enable users to view and modify the security status of their accounts, typically within settings. Examples include the *Where you're logged in* interface in Meta's Account Center [58], which lists devices associated with active sessions. These interfaces are incredibly valuable in clinical computer security consultations when survivors and consultants work together to detect suspicious account accesses [31, 38]. The Clinic to End Tech Abuse (CETA) additionally provides guides for navigating ASIs in common applications [78].

Although ASIs such as privacy and security checkups help people feel safer online [66], they can be both difficult to locate and to interpret [13, 21, 34]. Gallardo et al. [34] found that just two out of 18 users

could find the trusted device list in Apple's *Settings* ASI without help. Daffalla et al. [21] observed that survivors and consultants rely on incomplete details in ASIs to determine if an attack has occurred. If identifiers associated with a logged-in device (i.e., device model, operating system) are ambiguous, survivors find it difficult to determine whether it belongs to them or seems malicious. Multiple sessions on a single device may appear as distinct "devices," potentially causing confusion or alarm. Some ASIs only display *currently* active sessions rather than historical login actions, allowing an abuser to hide previous accesses made to the account. Historical logs may be time-limited or include only logins deemed unknown or risky by the application [21], which fail to consider that the abuser and survivor may have previously shared devices or a home network in interpersonal threat models [32, 33, 72]. In our study, we examine the extent to which data exports might be able to address several of these limitations.

2.3 Data Exports & Right of Access

Data protection laws such as Brazil's LGPD [60], India's DPDPA [40], and the European Union's GDPR [27] codify *right of access*, which allows users to request their personal data from organizations that collect or process it. The LGPD and GDPR also establish the *right to portability*, which mandates that organizations provide data in machine-readable formats so that users may transfer it to other services [27, 60]. The CCPA and CPRA (California, USA) contain a similar provision, *right to know* [11, 12]. Services have largely automated data requests through web forms and in-app pages available to logged-in users [62]. Facebook's request page, for example, is available at https://www.facebook.com/dyi [28]. The resulting *data export* is typically a ZIP archive containing structured or semi-structured user data, such as login history, message and browsing activity, and advertisements [9, 62]. Data exports may also be referred to as Subject Access Request Packages (SARPs) [47] or data takeouts [3]. Services often make *right of access* request features available beyond regions with explicit legal requirements, likely due to the difficulty of verifying users' native jurisdictions [42].

Compliance, security, & usability. Early research following the 2018 GDPR [27] observed high non-response rates [45, 62, 81, 85], deceptive patterns in request interfaces [52, 62], noncompliance with legal machine-readability requirements [45, 62, 85], and vulnerabilities during authentication [10, 54, 55]. More recent work suggests that the state of regulatory compliance and security has improved in the years since [45, 54]. However, even when users successfully download data exports, they encounter usability hurdles when viewing and understanding them. Veys et al. [82] found that users feel overwhelmed by the sheer size and disorganization of the data. Borem et al. [9] had participants annotate their own data exports with reactions of confusion, creepiness, interest, and surprise. To improve usability, Schufrin et al. [70] developed a visualization tool for users to upload and explore their data exports using a variety of views and filters.

Exports as datasets. Data exports themselves can be useful datasets for downstream tasks such as analyzing user activity. Ebbers et al. [25] reconstructed car usage from data exports from vehicle assistant apps. Onaolapo et al. [61] used data exports to examine in-the-wild usage of honeypot Facebook accounts. Razi et al. [63, 64] and Ali et al. [1, 2] used them to understand and detect harmful interpersonal interactions among adolescents on Instagram using metadata, media, and linguistic characteristics of messages. No prior work yet has looked at the feasibility of using data exports for investigating account compromise within the context of TFA.

Ethically acquiring ecologically valid data exports is often challenging [6]. Prior studies have either asked users to donate their data exports ("data donations") [1, 2, 63, 64] or analyzed exports from researcher-controlled accounts [25, 47, 61]. Concerns about participant privacy limit what can be learned from data donations, while researcher-controlled accounts may not fully capture authentic user behavior [6, 64]. Given the sensitivity of TFA, we chose the second method following the procedure outlined by Leschke et al. [47].

3 Method

We created a set of researcher-controlled accounts and simulated regular app usage to match the real-world behavior of hypothetical survivors and abusers in a TFA scenarios. We selected a total of six online services for our analysis. Two of these are run by operating system vendors: Apple iCloud [39] and Google [50, 51]. Given the ubiquity of iPhone and Android devices, most users are likely to have an account with at least one of these two platforms. The remaining four services are popular social media and messaging services: Facebook [57], Instagram [41], Snapchat [73], and Discord [23]. The first five services have been identified in prior research as commonly weaponized technology-facilitated abuse (TFA) scenarios [30, 32, 63, 83]. We included Discord because it has over 200 million active users and provides extensive account activity logs in its data exports [22].

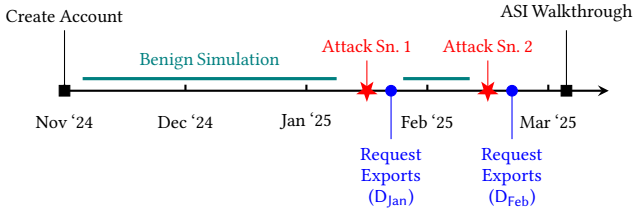

Figure 1: Simulation & Data Collection timeline. We first simulated benign usage activity for more than two months. Then, we simulated attacks from the perspective of the Alex (abuser) persona against Sam (survivor) on two separate days, shown as Attack Sessions 1 & 2 (Section 3.2). We requested data exports from Sam's account on each service within 8 days of each attack session, and refer to the two datasets as D_{Jan} and D_{Feb}.

3.1 Account Setup & Benign Simulation

We drew upon prior work by Leschke et al. [47] to create mobile research environments that minimize the collection of the research team's personal information. We created two pseudonymous user personas: **Alex**, who represents the abuser, and **Sam**, who represents the survivor. We designated each user persona a mobile phone: an iPhone 7 (iOS 15.7) for Alex and an iPhone XR (iOS 17.7.1) for Sam. Both phones remained on Wi-Fi and had no active SIM card. We used phones rather than PCs in order to access platforms through both mobile apps and the browser. iPhones were chosen to test iCloud

Attack Goal	Capability	Services	Attack Steps
A1. Account Surveillance	Alex knows Sam's credentials. Sam has MFA disabled.	All except iCloud.† Gmail accessed via browser; others via mobile apps.	1. Alex logs into Acc_{Sam} from $Phone_{Alex}$. 2. Alex browses emails / messages in Acc_{Sam} from $Phone_{Alex}$. 3. Alex logs out of Acc_{Sam}.
A2. Location Monitoring	Alex has authenticated access to $Phone_{Sam}$ and all logged-in applications.	Apple (FindMy), Google Maps (Location Share), Snapchat (Snap Maps).	1. Alex takes $Phone_{Sam}$, navigates to the service app and shares location with Acc_{Alex} on Acc_{Sam}. 2. Alex views Sam's location on Acc_{Alex} from the respective app on $Phone_{Alex}$
A3. Impersonation	Alex knows Sam's credentials. Sam has MFA disabled.	All except iCloud.† Gmail used via browser; in-app direct messaging for others.	1. Alex logs into Acc_{Sam} from $Phone_{Alex}$. 2. Alex sends a message impersonating Sam from Acc_{Sam}. 3. Alex deletes or unsends the message. 4. Alex logs out of Acc_{Sam}.
A4. Lockout & Control	Alex has authenticated access to $Phone_{Sam}$ and all logged-in applications.	All platforms. iCloud and Google accessed via browser; others via in-app password reset.	1. Alex attempts and fails to log into Acc_{Sam} from $Phone_{Alex}$. 2. Alex initiates password reset or recovery for Acc_{Sam}. 3. Alex resets the password using access to $Phone_{Sam}$. 4. Alex logs into Acc_{Sam} from $Phone_{Alex}$.

† We did not simulate these attacks within iCloud as it does not let users disable MFA and iMessage is not available through the browser.

Figure 2: Four attack simulation scenarios considered in our study. Alex (abuser) and Sam (survivor) are our two user personas. $Phone_{Alex}$ and $Phone_{Sam}$ are their primary devices, which they use to access Sam's account, or Acc_{Sam}.

features (e.g., Find My, iMessage) that are unavailable on Android devices. For each user persona, we first opened a new Google account via the designated phone's Safari browser and then used this email to create an Apple iCloud account. Next, we downloaded and set up accounts for the following mobile apps: Gmail [50] and Google Maps [51] (used existing Google account), Facebook [57], Instagram [41],[3] Snapchat [73], and Discord [23]. We created all accounts on Nov. 8th, 2024, although our Facebook account took an additional six days to be verified. Some account verification steps required a non-VoIP phone number or secondary email; if so, we used the first author's contact information, which was redacted during preprocessing (Section 3.4). We only set up multi-factor authentication (MFA) when it was required (iCloud), because prior work suggests survivors in TFA contexts often do not have it enabled [31, 38].

We created detailed simulation schedules for each account, which began with a period of benign usage to generate realistic data exports. During this period from account creation until mid-January 2025 (see Fig. 1), our Alex and Sam personas messaged each other, viewed and created posts, searched for content, and changed profile settings and authentication mechanisms. Following Leschke et al. [47], our research accounts only interacted with each other and large organizational accounts (i.e., universities, news sites). No messages or posted content contained identifying details. We recorded the start and end times of significant simulation steps in a shared simulation log document. After over two months of regular benign usage, we simulated four account compromise attacks, which we describe next.

3.2 Attack Simulations

To better understand how data exports represent SRI, we simulated account compromise attacks by an interpersonal adversary, drawing upon common experiences of TFA [31, 33, 56].

[3] Although both Instagram and Facebook are subsidiaries of Meta, we treat them as separate due to distinct in-app request interfaces and different data export formats.

Threat model. Many, if not most, adversaries in TFA contexts are *UI-bound* [33], conducting attacks only through application user interfaces while aided by physical or interpersonal means. We consider two attacker capabilities:

(1) *Known Credentials:* Prior work indicates that survivors share passwords with abusers, sometimes willingly and sometimes by coercion [19, 24, 31, 33]. An abuser may know private information about the survivor, such as children's or pets' names, that makes it possible to guess credentials [32].

(2) *Authenticated Device Access:* Even if account credentials are not shared, abusers can exploit physical proximity to survivors' devices or knowledge of device PINs to bypass account authentication mechanisms [32]. Notifications containing 2FA and account recovery codes are often visible on locked devices.

Attack simulation design. We designed four attacks based on categories of TFA threats identified by Thomas et al. [77] that require privileged access to a survivor's accounts and devices: surveillance, impersonation (specifically, "hijacked communication"), and lockout & control. Because surveillance encompasses a range of threats, we consider location monitoring separately from surveillance over general account activity like posts and messages.

For each attack, we defined a set of step-by-step actions (see Fig. 2). One author simulated the role of **Alex** (abuser) carrying out attacks against devices and accounts belonging to **Sam** (survivor). We conducted all four attacks consecutively in a single *attack session*. Because several ASIs only show security logs for 28–30 days [21], we performed two such attack sessions 30 days apart, on January 21st and February 20th. Fig. 1 shows a timeline of the entire simulation procedure.

3.3 Collecting Data Exports

Within eight days of each attack simulation, we requested data exports for all six of Sam's accounts using the research phone ($Phone_{Sam}$). We did not request Alex's data exports, as their data would contain

no relevant security-related information (SRI) for investigating attacks against Sam. All services except Discord allowed us to select specific subsets of data, file size limits, or preferred formats. We chose all available subsets, the largest file size (10–25GB), and CSV/JSON over HTML when possible. Most services delivered download links within 2–4 days, though Apple delayed one file set by 17 days due to temporary service unavailability. Five of the six services provided a single ZIP archive; Apple instead delivered multiple archives. We refer to the datasets requested after the January and February attack simulations as D_{Jan} and D_{Feb}, respectively. Each contains six data exports, one per service, recursively unzipped and preserving original file structures.

Basic statistics for each data export appear in Fig. 3. As our research accounts were new, most social media data exports were small (<10 MB). Apple's and Google's were larger (up to 827 MB), as they included cloud-uploaded media files; for example, screen recordings from our ASI walkthrough (see Section 4) were auto-uploaded to iCloud Photos. Overall, D_{Feb} contained 40 more files (10.3%) than D_{Jan}. Notably, Snapchat's data export decreased in size as the service appears to delete chat media after 30 days.

Platform	Export	*Original*		*After Preprocessing*		
		MB	Files	MB	Files	Elements
Apple	Jan.	47	94	0.8	33	2,931
	Feb.	827	130	0.9	28	3,348
Discord	Jan.	6.2	10	6.2	10	3,695
	Feb.	9.2	10	9.2	10	5,579
Facebook	Jan.	0.3	112	0.2	97	465
	Feb.	0.3	112	0.2	97	708
Google	Jan.	22	91	9.5	29	2,698
	Feb.	113	101	10.6	32	2,633
Instagram	Jan.	0.2	41	<0.1	36	188
	Feb.	0.2	42	<0.1	37	260
Snapchat	Jan.	12	41	<0.1	17	136
	Feb.	0.3	34	<0.1	15	138

Figure 3: Data export sizes. The *Original* column shows export sizes exactly as they were downloaded. Google's and Apple's exports are large because we used their Drive and Photos features to collect data exports and screen-recordings. The *After Preprocessing* column shows export sizes file filtering (Section 3.4) as well as the number of *data elements* parsed from each data export (excluding null or empty elements).

3.4 Preprocessing

After obtaining files, we prepared them for further analysis to identify and systematize the information within them, especially since not all of it applied to our research objective. First, we pseudonymized both datasets to protect any remaining identifying information. Next, we filtered out files if they lacked machine-readable text or contained only data about features outside the scope of our simulation. To enable consistent analysis across all services, we developed a unified *entity-attribute-value* (EAV) model and parser for data exports, allowing us to query and sort the data chronologically.

Pseudonymization. Although we isolated our research devices and accounts to the extent possible in a UI-bound simulation [47], additional steps were required to protect the authors' privacy. For instance, Apple and Facebook required a non-VoIP phone number to validate new accounts. Snapchat provided precise coordinates in its location history file, and all services recorded IP addresses. We replaced any potential personally identifiable information — such as phone numbers, IP addresses, names of friend suggestions, and any internal identification strings — with syntactically valid pseudonymous values. For example, each unique phone number was replaced with `"000-00-000x"`, where `"x"` was different for each. Similarly, IP addresses were masked like `"0.0.0.x"`.

Filtering Files. We removed files that did not contain machine-readable text data, like images and videos. Additionally, we removed files related to features we did not interact with during the simulation phase, including IoT integrations, streaming activity, payments, and educational features (full list provided in parser documentation). While these files may be useful in investigating other TFA incidents — namely, abuse via IoT devices [14, 74, 75] and financial services [7] — we limited our scope to mechanisms of account compromise (see Section 3.6). Fig. 3 shows the sizes of both sets of data exports before and after filtering. All remaining files were HTML, JSON, and CSV, with the exception of a README file from Discord and an MBOX file from Google.

Data Model. Because our data exports are sparse and differ widely in format, we broke files into *data elements*, individual chunks of data representing events or objects within a machine-readable file [70]. Then, we parsed data elements in JSON, CSV, and HTML files into an *entity-attribute-value* (EAV) model [53]. Each *entity* is a data element, such as a CSV row or a JSON dictionary within a list. We use *entity* and *element* interchangeably. *Attributes* are properties of the entity or element, usually column headers or JSON keys. *Values* are the corresponding cells or dictionary values for each attribute. Each element additionally has a reference to its service, file path, and a unique ID assigned by the parser. We provide an example in Fig. 4.

Parser Implementation. File formats, even from the same service, tend to vary widely [9] and no services provided machine-readable schema definitions. Due to this variation, we designed our parser to handle file *types* rather than individual files. For CSV files, each row is an element, with column headers as attributes and corresponding cells as values. Most JSON files we encountered were lists of one-dimensional dictionaries; in this case, we parsed each dictionary as an element. We flattened nested keys using string concatenation with a delimiter. If a list of dictionaries was nested within another, we flattened the hierarchical data structure. An example of this transformation is provided in the parser documentation on GitHub. Since Google provided some critical SRI only in HTML format, our parser processed HTML tables like CSV files and nested HTML elements like JSON files. The number of data elements parsed from each data export is given in the right-most column of Fig. 3.

The parser additionally extracted date and time fields, if present, from the attribute column using regular expressions. If an element contained multiple timestamps, we used the most recent.

File Anomalies. Consistent with previous findings [62, 85], we encountered machine-readability issues such as poorly-escaped characters and improperly formatted files. For example, Apple provided

data elements

file	ID	ts	attribute	value
Google Account/ sam.researcher24. SubscriberInfo.html	2675	1/21/2025 17:42:42	Timestamp	*2025-01-21 17:42:42 Z*
			IP Address	*0.0.0.10*
			Activity Type	*Logout*
			Interactive	*TRUE*
			Geo	
			Raw User Agents	*Mozilla/5.0 (iPhone; CPU iPhone OS 12_1_3 like Mac OS X) ....*
			Challenges (timestamp, outcome, dusi)	
Google Account/ sam.researcher24. SubscriberInfo.html	2676	1/21/2025 17:41:54	Timestamp	*2025-01-21 17:41:54 Z*
			IP Address	*0.0.0.10*
			Activity Type	*Login*
			Interactive	*TRUE*
			Geo	
			Raw User Agents	*Mozilla/5.0 (iPhone; CPU iPhone OS 15_7 like Mac OS X) ....*
			Challenges (timestamp, outcome, dusi)	*2025-01-21 17:41:54 Z: Challenge Passed, -*

Figure 4: Example of data from Sam's Google export parsed into our *entity-attribute-value* data model. This file in question is an HTML file containing a table of login and logout events. We parse each of the outer rows in this table into an *element* and assign it an integer ID. Column headers and their corresponding cell values become *attributes* and *values*, respectively. This format unifies sparse data across multiple files in a single table, allowing us to perform structured queries and sort by time.

several malformed CSVs containing multiple CSV segments concatenated in the same file. In this case, our parser processed each segment separately. In Instagram's and Facebook's data exports, we observed several files in which JSON keys were literal references to data structures (e.g., `"vec"` and `"dict"`). We adjusted our parser to read applicable file content directly into the described data structures.

Grouping elements. As noted in Figure 3, we ended up with thousands of individual data elements for some services, far too many to manually label. In their data export visualization tool, Schufrin et al. [70] categorized data based on substrings within the filename. However, we observed that some platforms (notably Discord) provided a large quantity of heterogeneous data in a single log file. To account for this, we identified a set of *category attributes* for each service that explicitly describe the category or type of a given element. In the elements shown in Fig. 4, `"Activity Type"` would be a category attribute. We generated our list of category attributes by querying a set of seed terms and manually removing false positives (full list available in the parser documentation).

Then, we grouped two elements if they (1) originate from the same file, (2) have the same set of attributes, and (3) have the same value for each category attribute. We refer to these three requirements as a group's *characteristic features*, and we assigned each group an ID. For example, the two elements shown in Fig. 4 have the same file and attributes, but they do not have the same value for `"Activity Type"`. The first element would be grouped with other `"Activity Type: Logout"` elements, and the second would be grouped with `"Activity Type: Login"` elements.

There is a possibility that we failed to group two similar elements or incorrectly grouped two elements. The former would have increased our manual coding efforts and would not have impacted the accuracy of our analysis, while the latter could lead to missed elements. Therefore, we manually verified a random sample of elements from each group; we did not find any erroneous grouping.

Labeling groups using a qualitative content analysis. We labeled each group using a set of codes, or labels, that we derived using qualitative content analysis of the data export content (more details in Section 4). We created a mapping of each group's characteristic features to a set of labels to analyze future data exports.

Processing D_{Feb} with our pipeline developed for D_{Jan}. We initially developed our data analysis pipeline using D_{Jan}. To validate its generalizability, we applied the same on D_{Feb}. We filtered, parsed, and grouped elements present in D_{Feb} by the same procedure as above. For each group in D_{Feb}, we applied labels from D_{Jan} based on shared characteristic features. Due to schema changes between D_{Jan} and D_{Feb}, we observed that our parser and the grouping heuristic were unable to map labels to every element in D_{Feb}. For example, Facebook and Instagram removed two keys (`"ent_name"` and `"ent_field_name"`) from one of their common JSON file formats. Apple added a total of 6 new segments to several of its malformed CSV files discussed before.

After accounting for such changes, we found that we could label all but 45 groups of elements in D_{Feb} using our pipeline. Among these, 13 were due to new files in D_{Feb} (provided by all services except Discord), and 32 were due to new category attribute values. We manually applied labels to the 45 new groups in D_{Feb}, and found no new types of information that we had not yet encountered in D_{Jan}. We assigned labels based on exact matches of group characteristic features; however, future work could explore fuzzy matching or LLM techniques to accommodate minor variations.

Final Datasets. After completing our data export processing pipeline, we generated a database table (like Fig. 4 with additional columns for labels) for each service containing all elements from its data export. Each element has references to a unique ID, service name, source file path, timestamp, and a set of attribute-value pairs; in addition to label(s) generated during qualitative coding (Section 4). This structure enables efficient querying and analysis using standard database tools, which we used in Section 5 and Section 6.

3.5 Positionality & Ethical Considerations

The authors have expertise in digital safety and violence prevention, with three providing direct technical support to TFA survivors using trauma-informed care [69]. This advocacy shaped our research questions, attack simulations, and the sensitive framing of our findings.

As discussed in Section 2, data exports often contain sensitive data like emails, chat logs, and location history that may be unsafe or uncomfortable for survivors to share. It was initially unclear which files held relevant information or how to safely redact them, so the authors reviewed their own data for familiarity but did not contribute them for analysis. Instead, all data exports were generated using researcher-controlled accounts with simulated behavior, with selected data redacted to protect the authors' privacy. As no human subjects were involved, this study did not require IRB review.

3.6 Limitations

One limitation arises from our choice to use iPhones in our simulations. Online platforms may collect different quantities of data from the origin device depending on its manufacturer or operating system; e.g., Google's takeout may contain more hardware identifiers for Android phones. We argue this was a necessary design choice to test both iCloud and Google features (see Section 3.1), but future work should aim to run similar tests with a wider range of devices.

We also note that not all TFA involves account compromise. For example, an abuser can harass a survivor by sharing non-consensual intimate imagery (NCII) without accessing the survivor's account. In such cases, data exports offer little value, as no relevant information would appear in the survivor's account export. However, in TFA cases where the abuser does access the survivor's account (e.g., sending harassing messages, posting harmful content, or changing settings), data exports can provide critical evidence, including message logs, posts, and session metadata. We focused specifically on account compromise, as these cases are both common among survivors and difficult for advocates to investigate using existing tools.

4 Security-Related Information in Data Exports

In this section, we address **RQ1**: *What types of information present in data exports are relevant to investigating account compromise by an interpersonal abuser?* Data elements in exports generally fall into two broad categories: (1) information describing the current *state* of the account at the time the data export was generated, and (2) historical *events* or actions initiated by the user or application. For instance, a list of current privacy settings or authenticated devices reflects the account's *state*, while logs of setting changes or time-stamped login records would be *events*. We argue data exports are particularly valuable for investigating account compromise because they contain lists of past security-critical *events* that may not be otherwise accessible to the survivor through ASIs.

Method: Qualitative Content Analysis. Consistent with prior work [9, 82], we observed that naming conventions in data exports are semantically ambiguous and inconsistent. Accordingly, we treated data export schemas as qualitative data sources. While earlier efforts have relied on "top-down" parsing of data into predefined behavioral categories via techniques such as regex matching [61, 70], we found that these categories lack the granularity needed to analyze nuanced security-related information (SRI) in account compromise scenarios. Therefore, we adopted a "bottom-up" qualitative content analysis approach [26, 29, 71] to construct a taxonomy of SRI directly from data export schemas.

Our unit of analysis was each *data element* from the data model described in Section 3.4. As some services yield thousands of elements (see Fig. 3), we reduced manual workload by coding a representative instance per group of elements.

We used both inductive and deductive coding [44, 67, 68], using a taxonomy of adversarial account behavior by Onaolapo et al. [61] as our initial codebook. The first author coded one element per group in D_{Jan}, proposing new codes when existing ones were inadequate for our use case. For example, we refined "Hijacker" into more granular categories: *account creation*, *authentication*, (active) *interaction*, (passive) *viewing*, and *notification*. We also flagged "background activity" not explicitly driven by user interaction.

The team iteratively refined the codebook by adjusting definitions and categories for consistency. After revisions, the first author re-coded D_{Jan}, and another author independently coded a random sample of 50 entries. We resolved eight discrepancies, primarily due to ambiguous definitions, through group discussion. Our codebook is available in the Appendix.

Results. We identify two categories security-related information relevant for account compromise in data exports: (a) data related to user *authentication*, such as login and logout events, and (b) data concerning security *settings*, including changes to email addresses, passwords, recovery methods, and multi-factor authentication. Furthermore, we systematize (c) logs of app activity, notifications, other profile settings, and user interactions that — while not directly security-related — may have important implications for understanding what an abuser accomplished in a compromised account.

(a) Authentication logs. We observed a range of authentication-related information in data exports, including records of logins in all services, logouts on all platforms but Apple and Snapchat, and even failed login attempts (Google).

Authentication information was often scattered across multiple files, making analysis difficult. Google's login records were split across folders, with some hidden in unintuitively named files like `"<username>.SubscriberInfo.html."` Some platforms provided a primary folder for this data with a clear, security-related name, but key details frequently appeared elsewhere. For example, Meta provided a primary `"security_and_login_information"` folder, but in-depth device identifiers were in the `"personal_information"` folder. Redundancy of SRI across multiple files was common. Many files in Meta's primary security folder were simply subsets or aggregations of other files. Apple provided daily, weekly, and monthly summaries, and Discord duplicates the same events across four files. This overlap can easily mislead survivors or advocates, causing them to mistake one login for multiple unauthorized attempts.

Authentication logs also vary in granularity and often blend user-driven actions (e.g., logins) with system-driven ones (e.g., token refreshes). Facebook's `"account_activity.json"` file, for instance, mixes these events without a clear distinction. Apple stores its main logs in `"Apple ID account and device information,"` but additional analytics are buried in vast CSV files under folders with opaque names like `"Other Data Part 4 of 5."` Similarly, this may mislead or alarm users.

(b) Security settings. The majority of services provided records of *changes* to critical security settings such as passwords and emails. All data exports contained logs of changes to the primary recovery email associated with the account (including both the old and new addresses), even after the recovery email has been removed from the account. All platforms except Discord recorded changes to passwords, although not the password itself (for good reasons). Discord's data export did not explicitly log password changes, yet it did record email notifications sent to the user about password recovery attempts, as well as when those emails are opened. In cases of account lockout or surveillance, understanding the history of password, email, and phone number changes can help survivors determine who had access to their accounts and when.

(c) Activity and notification logs. Data describing other user-driven activity (e.g., interactions with other users, browsing and page-view history, and logged notifications) can be incredibly helpful to understand *what* an abuser accomplished after obtaining account access. All social media services provided logs of posts, comments, and messages (both sent and received) from the account in question. Apple's data export did not contain iMessage data, which is end-to-end encrypted. Google provided sent and received emails in an MBOX file. Google, Facebook, and Instagram logged recent search history. The degree to which services logged *passive* use, such as opening apps or viewing content, varied. Discord recorded intense click analytics whenever the user opened chat channels or viewed settings pages. Facebook and Instagram both logged URLs of media content viewed from the account. Google provided thousands of lines of app access data in `"Activities - A list of Google services accessed by.csv,"` essentially periodic pings whenever a user navigates to a page or loads content. Facebook similarly recorded timestamps for recent intervals of activity. Each service recorded at least a subset of email *notifications* sent from the service to the user (in Gmail's case, these are available in the MBOX file). While most are marketing emails, some (like Discord, as mentioned above) respond to security-critical events. Regardless, notification data may help survivors contextualize whose contact information was associated with their account at what times.

5 Comparing Data Exports with ASIs

In this section, we answer **RQ2**: *How does security-related information (SRI) within data exports compare with information presented in account security interfaces?* Account security interfaces (ASIs) are essential for survivors (and the advocates who support them) to identify if their accounts have been compromised [21]. However, prior research [21] has noted that some ASIs are seriously limited in their ability to protect survivor's safety. Many leave out critical login information, such as originating IP addresses (Apple, Discord, Google, and Instagram) or exact times (Apple, Snapchat, and Discord). Some ASIs display only *currently* active sessions instead of past logins, which can let abusers conceal previous account access [21]. Descriptions of logged-in devices are often limited to coarse-grained information like `"iPhone"` or `"Samsung,"` which can make it quite difficult for survivors to identify which devices are benign. We evaluate whether data exports can address several of these limitations by comparing security-related information in data exports against ASIs.

Method: UI Walkthroughs. We systematically analyzed the contents of the ASIs via UI walkthroughs [43, 49] within 2–3 weeks of our February attack simulation (see Fig. 1), ensuring we could capture time-limited data before it disappeared [21]. We followed instructional guides published by the Clinic to End Tech Abuse (CETA) [78] for locating ASIs and changing critical settings. These guides are designed to help survivors check their accounts, but they are also often used by tech clinic consultants [18, 80]. As no guides existed for Discord, we navigated through all settings related to authentication, user profile, and sharing, such as *Settings* > *Device*, *Account*, and *Privacy & Safety*. We examined the mobile app interfaces of all six applications and extended our analysis to also include the browser-based ASIs for Google and Apple.

During our walkthrough, we recorded the screen and took detailed notes about all SRI (authentication and security-critical settings, as described in Section 4) for each ASI page visited. We additionally noted the retention period and device identifiers associated with security events. We build our initial list of device identifiers from [21] and add additional types as we encounter them.

Key findings. We identify three core differences between the SRI available in the D_{Feb} dataset and ASIs: (1) data exports offer higher-resolution identifiers of logged-in devices than ASIs; (2) some types of SRI are available exclusively through data exports; and (3) data exports retain longer histories of certain security events compared to what is shown in ASIs. We discuss in greater detail below.

5.1 Device Identifiers

We found that data exports contained more identifying information about *currently logged-in devices and active sessions* (e.g., hardware, software, network, and location attributes) than their ASI counterparts for all six services. Moreover, there was no case in which an ASI contained identifying information about a device that was not either explicitly present or easily inferred from data export contents. Fig. 5 presents a side-by-side comparison of identifiers observed in ASIs and data exports.

Device Hardware & Software Information. Because iOS does not typically allow 3rd party apps access to unique hardware identifiers such as device serial numbers and IMEIs, only Apple presents these. All ASIs provided high-level device models ("iPhone" or "Android") for all logged-in devices and current sessions, but this may be insufficient information for survivors if their abuser uses a device of the same type (like the two iPhones in our simulation). In this case, the model version (iPhone XR vs. 7) would provide more utility. Fig. 6 shows the asymmetry between Discord's ASI and data export; the ASI shows just "iOS," while the export supplies the device's model and version. Although the value `"iPhone11,8"` seems to be a mistake, this is a *mobile device code* referring to the iPhone XR.[4] However, the model version was not always available, for example, in the *Mobile Safari* session in Fig. 6. We observed that many services included this identifier for mobile app sessions, but omitted it for browser sessions. This may be due to recent privacy enhancements implemented by Apple on the Safari browser to reduce information in user agent strings [46]. Still, device identifiers from browser sessions, even when limited, can help survivors recall whether a login was legitimate. For instance, Google's logs of browser-based accesses still contained iOS versions (15.7 and 17.7.1). This allowed us to distinguish between Alex's and Sam's iPhones, which we could not do with the ASI alone.

Multiple sessions on the same device. Confusingly, multiple sessions from a single device often showed up in ASIs as separate "devices." Fig. 6 shows an active browser session as a different device even though we access Discord only through its mobile app. We hypothesize this is because some in-app settings redirect to the in-app browser. We found generally that data exports offered more clarity between browser and app-based sessions. Similarly, Google's ASI

[4] See additional examples here: https://gist.github.com/adamawolf/3048717 [84]

Identifier Type		Apple	Discord	Facebook	Google	Instagram	Snapchat
Hardware	Serial number	●	×	×	×	×	×
	IMEI	●	×	×	×	×	×
	Model *(iPhone)*	●	◐	●	●	●	●
	Model version *(XR)*	●	◐	●	●	●	●
Software	User agent	×	×	◐	◐	◐	◐
	OS *(iOS)*	●	●	●	●	◐ D	●
	OS version *(15.7)*	●	◐	◐	◐	◐ D	◐
	Locale/language	◐	◐	◐	◐	◐	◐
	Browser type	×	●	◐	● D	◐ D	●
Network & Location	IP address	◐	◐	●	◐	◐	●
	Country	◐	● D	●	●	●	● A
	State	◐ D	● D	●	● D	● D	● D
	City	◐ D	● D	●	● D	● D	● D
Other	Device name	◐	×	×	×	×	×
	Session cookie/ID	×	◐	◐	×	◐	×
	Internal device fingerprint	×	×	◐	×	×	×
	Time of login	◐	◐	●	●	●	◐

Figure 5: Device and network identifiers in data exports and ASIs describing currently logged-in devices. We label identifiers as present in both ASIs and data exports (●), only in data exports (◐), or in neither (×). While "Time of Login" is not a standard device identifier, prior work [21] shows that survivors often use timestamps to spot suspicious activity. Fields that can be easily inferred (e.g., location from IP or device details from user-agent strings) are marked as present. We use the superscripts "D" for inferred fields in data exports and "A" for those in ASIs.

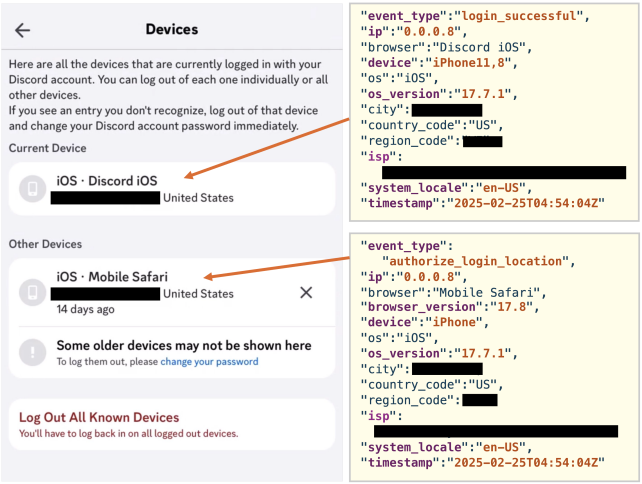

Figure 6: Identifiers in Discord's ASI vs. data export The screenshot on the left shows the *Settings* > *Devices* page in Discord's iOS app for Sam's account when only Sam is logged-in [23]. The code blocks on the right are two (truncated) dictionary entries in the export mapping to the sessions on Sam's iPhone XR. The bold fields show device and network identifiers not present in the ASI.

showed 7 active "devices," but Google's trusted device file (`"Devices - A list of devices (i.e. Nest, Pixel, iPh.csv"`) seems to resolve this redundancy, logging only two phones.

IP address & location. All data exports provided current or last-used IP addresses for current sessions, but only Facebook and Snapchat displayed them within their ASIs. Most ASIs instead displayed the login location, which appeared to be inferred from the client IP address. Location identifiers typically consist of country, state or region, and city. However, there are cases where an IP address itself would be more useful than general location data, for example, if the survivor and abuser live in the same city. In a clinical setting, a trained consultant could use the IP address to identify other actions originating from the abuser's device or network. We experiment with the feasibility of this in Section 6.

Robustness. Daffalla et al. [21] demonstrated that unauthorized accesses to accounts can be obscured by spoofing IP addresses and user agent strings. We did not test if data exports are more resistant to spoofing, but we have no evidence to suggest they are more so than ASIs. This likely extends to derived fields such as inferred location and device model/OS, with the possible exception of Apple's and Google's hardware identifiers for Apple devices and Androids, respectively.

5.2 Historical SRI

Many ASIs show only active sessions or currently logged-in devices. Daffalla et al. [21] demonstrated that *access hiding attacks* are possible on some services if an attacker terminates a session and therefore conceals their login from the ASI. We found that data exports provided significantly more information compared to ASIs about historical authentication events and changes to critical security settings, as shown in Fig. 7.

Sessions & Logins. In [21], access hiding attacks were successful on Apple/iCloud, as well as in Google's *Recent security history* interface, if Google failed to classify the login as risky or unknown. We confirmed this in our ASI walkthrough and found this also happens to some extent on all services except Facebook. Instagram also appears to display only logins classified as "new." Its *Recent Emails* tab

Platforms		Active Sessn.	Hist. Logins	Logouts	Pwd. Change	Email Change
Apple	DE	●	●	×	●	●
	ASI	○	×	×	○	×
Discord	DE	●	●	●	×	●
	ASI	●	×	×	×	×
Facebook	DE	●	●	●	●	●
	ASI	●	●	●	○	×
Google	DE	●	●	●	●	●
	ASI	●	○	×	○	○
Instagram	DE	●	●	●	●	●
	ASI	●	○	×	●	●
Snapchat	DE	●	●	×	●	●
	ASI	●	○	×	×	×

Figure 7: Active and historical security events shown in data exports (DE) and ASIs. The figure indicates whether all (●), a subset (○), or none (×) of the simulated events (i.e. all active sessions or all historical password changes) appeared in the data exports or ASIs.

claims to show logins that triggered email notifications in the past 14 days. However, Alex's logins during our simulations did not appear in this tab, even within the specified timeframe. We suspect this is because Alex's iPhone 7 had previously accessed the account during the January simulation and was therefore not considered new. This is concerning, as known logins are not always *safe* logins from an interpersonal technology abuse perspective.

Discord, Instagram, Facebook, and Google provided both past logins and logouts in their data exports. Google also recorded failed login attempts in `"Google Account/<username>.SubscriberInfo.html."` Discord recorded the login method, which in our case was always `"password"` except immediately after account registration. Apple provided multiple files with data about authenticated devices, but only one (`"Apple ID account and device information/Apple ID SignOn Information.csv"`) showed time-stamped login histories from both Apple devices and browser-based accesses. However, this file contained far fewer device identifiers (IP address only) than other files with authentication data.

Password & Email Changes. In cases of account lockout or takeover, knowledge of the previous password and email changes can help survivors understand who has had account access at what times [77]. Several ASIs displayed the date of the last password change, but only Instagram provided the full password and email change history in its interface. Google showed password/email changes from the last 30 days in its *Recent security history* interface.

In our data exports, all services provided timestamped logs of email and password changes, along with the new and previous email values. Several ASIs (Apple, Facebook, Google) showed the date of the most recent password change, but not all historical changes. Instagram's ASI provided comprehensive records of both.

5.3 Log Lifetimes

As noted by Daffalla et al. [21], some ASIs only show SRI within a limited time window. For example, Google's *Recent Security Activity* ASI states that it displays security events from the last 28 days, while Gmail's web-based *Activity on this account* interface shows only the 10 most recent sessions.

In contrast, we found that all data exports in both D_{Jan} and D_{Feb} include login history dating back to account creation. For D_{Feb}, this spans roughly 102–125 days, depending on the service and the delay between requesting and receiving the data.[5] We cannot confirm whether data exports always retain login data beyond this window. Some export files containing generic user activity had obvious retention limits. For example, Google's large `"Activities.."` file[6] contained approximately 30 days of history: the earliest entry in D_{Jan} is from Dec 30, 2024, and in D_{Feb} it is Jan 28, 2025. Snapchat similarly limits chat metadata to about a month in `"chat_history.json"`, which is consistent with its ASI.

6 Mapping Attacks in data exports

In our final analysis, we address **RQ3:** *To what extent can we map data exports to abuser actions during account compromise attacks?* Prior research [31, 65] suggests that once users suspect compromise, they often seek answers about how access was gained and what actions were taken. As we demonstrated previously in Section 5, ASIs alone often fail to provide a complete picture of this.

We explored how actions taken by Alex's persona using Sam's accounts were represented in our datasets. Using ground truth from our simulation logs (precise timestamps and device identifiers), we mapped specific steps from each of the four attacks we simulated (**A1–4** in Section 3.2) to data elements in the February dataset (D_{Feb}).

Identifying the time and method of authentication in **A1**, **A3**, and **A4** was relatively straightforward across platforms. However, our ability to reconstruct post-authentication activity varied depending on the service and type of action taken. Discord's detailed click analytics and Google's access logs allow us to link actions to specific devices or sessions. In contrast, platforms like Instagram offer limited visibility. Notably, we find no evidence of ongoing location-sharing activity in any data export, despite its likely status as personal information under right of access.

While real-world scenarios lack this level of certainty, our controlled setting enables exploratory analysis. We discuss the practical challenges and opportunities for using data exports in clinical computer security consultations [38] in Section 7.

Method. For each data export in D_{Feb} and attack scenario, we queried for attack-related data elements (1) by time and then (2) by known identifiers. First, we sorted the preprocessed data export by time and used our attack simulation log from February 21st to filter elements within 5 minutes of the attack window. From this, we analyzed a selection of *Authentication*, *View*, *Interaction*, *Settings*, and *Notification* events (see the Appendix for precise definitions). Second, we queried Alex's (attacker) known device and network identifiers on the entirety of D_{Feb} to account for possible incorrect or missing timestamps: `"iPhone 7"`, `"iPhone9,3"` or `"iPhone9_3"`

[5] Some services include data between the submission of data export request and receiving the data in their data exports. For example, Apple's D_{Feb} contained data from March even though we requested the data export on Feb 24, 2025.

[6] Full file path: `"Access Log Activity/Activities - A list of Google services accessed by.csv"`)

(device model code obtained from [84]), and "0.0.0.23" (masked IP address of the attacker during the attack simulation).

For each of the four attacks we simulated (Fig. 2), we discuss the feasibility and comprehensiveness of mapping the attack based on the data exports below.

A1. Direct Account Surveillance. This attack involved Alex (abuser) logging into Sam's (survivor) account using $\text{Phone}_{\text{Alex}}$ and opening Sam's messages/emails. We successfully queried Alex's login in all data exports, as each recorded historical logins (Fig. 7) with an IP address and user agent/device model at minimum (Fig. 5). When Alex accessed Sam's Gmail account through the browser, Google logged just "iPhone" in the user agent string instead of the full model name. The remaining services provided the device's model version (e.g., "iPhone 7"). Instagram, Facebook, and Google recorded logout events at the expected time with the same set of device identifiers as the login events. Only Discord indicated that Alex opened Sam's messages due to its granular logs of user click behavior. Although Discord's data export contained events labeled "session_end" elsewhere in the dataset, no such event was recorded at the time of our simulated logout.

A2. Location Sharing Surveillance. In the location sharing attack, Alex shared Sam's location on Snapchat, Google Maps, and iCloud Find My Friends with their own account, taking advantage of their authenticated access to $\text{Phone}_{\text{Sam}}$. We were unable to find records of Sam sharing location with Alex within any services' data exports. Although Snapchat, Google Maps, and Apple all collected device location data passively, their data exports do not indicate whether that information has been shared with anyone and, if so, with whom it was shared. We failed to find any indication that services retain past or terminated location-sharing activity. Users can still access their *current* location-sharing status through the app interface, however.

A3. Impersonation Attack. In this attack, Alex logged into Sam's account (Acc_{Sam}) from their own phone and actively sent messages as if they were Sam. Alex then deleted the messages, wiping their trace from Acc_{Sam}. Among the services we tested under this attack scenario, login and logout events appeared similarly to those in **A1**. Each provided logs of non-deleted chats, and Google provided trashed (but not permanently deleted) emails in the MBOX file. Among social media services, only Snapchat and Discord's exports recorded message deletion, though neither provided the deleted content. Discord recorded two events with "event_type: send_message" or "message_deleted" at the expected timestamps with attributes "channel_id"[7] and "message_id." Within the "messages" directory, we used "channel_id" to determine which chat channel contained the deleted message, but "message_id" did not map to any existing message content. Snapchat's file "chat_history.json" contained a record of a message sent from Sam to Alex with status "STATUSERASEDMESSAGE" and a null "Content" field. Snapchat indicated when this message was sent, but not when it was deleted — just *that* it was deleted.

For Discord and Google, a significant amount of user activity unrelated to authentication (e.g., searching, sending messages) could be traced to its originating login through multiple files. Discord's message events contained an IP address and detailed device model/OS data that we could attribute directly to $\text{Phone}_{\text{Alex}}$ and its associated login event. While Google's email records did not contain device/network identifiers, we used the email's timestamp to identify an app access event in the "Activities..." file associated with Alex's IP address, which then mapped to more detailed identifiers in "<username>.SubscriberInfo.html". This was possible, even though two devices were logged in simultaneously, due to the precision of the app access timestamps.

A4. Account Lockout Attack. In this attack, we simulated an attacker requesting a password reset on Acc_{Sam} following multiple failed login attempts. While we did not change other recovery methods (e.g., email or phone number) during **A4**, we verified email changes were captured in data exports during our benign activity simulations, as shown in Fig. 7. Across all services, we were able to query either the password reset initiation or the actual password change, often along with device identifiers such as user agent strings and IP addresses. Only Google's data export includes prior failed login attempts, available in "<username>.SubscriberInfo.html." Facebook and Google included both IP addresses and user agent strings related to the password change, while Apple, Instagram, and Snapchat recorded only timestamps. Facebook also logged password reset requests. Although Apple did not show all logins within its ASI (only currently active sessions), its data export captured the attacker's login from the expected IP address, lining up with our simulation timeline. While Discord did not explicitly record password changes, the service logged server-generated emails, such as password reset messages. These contained timestamps and recipient email(s), which could be useful for understanding how the abuser gained access.

7 Discussion: Operationalizing Data Exports for TFA Investigation

In this study, we demonstrate that data exports contain security-related information that can support investigations of account compromise. Specifically, in the context of TFA, we show that data exports provide more complete and historically rich logs of account activity than account security interfaces (ASIs), and they can be instrumental in tracing past attacks. While our findings highlight the feasibility of utilizing data exports for this purpose, several challenges must be addressed to operationalize their use in practice.

Identifying Attacks in Practice. In our simulated scenarios, we knew the ground truth about when and how account compromise occurred. In real-world cases, however, it is often difficult to confidently determine whether unauthorized access has taken place, especially without contextual cues. Even with technical skills, identifying adversarial activity (e.g., via IP address anomalies) depends on factors like geographic proximity of the abuser and home network configurations. Given the size and sensitivity of data exports, we recommend that consultants continue using ASIs for initial triage. If approximate timestamps or device identifiers are available, a targeted review of the survivor's data export may be considered when it is safe and appropriate.

Machine Readability & Standardization. For data exports to support investigations, they must be machine-readable and consistent. We encountered issues such as Apple's use of nonstandard CSVs with

[7] A "channel" is a direct message with another user in Discord.

poor character-escaping and Facebook's rapidly changing schema. Prior studies note similar challenges [62, 85]. While smarter parsers can help, the lack of standardization limits the long-term viability of tools built on top of data exports. Future *right of access* regulations should provide guidelines for schema versioning, changelogs, and machine-readable definitions to support maintenance of tools such as ours and promote research reproducibility.

Data Collection & Transparency. We recognize the tension between user privacy and security when it comes to long retention periods, detailed activity logs, and granular device identifiers, all of which we observed in data exports (Section 5). While we do not encourage services to collect *more* information from users, the fact that services retain more data than they make easily accessible demonstrates a lack of transparency. We urge services and developers to make the critical security information *that they already collect* readily available to those who need it.

Survivor Privacy in TFA Investigations. Data exports often include extraneous or highly sensitive data unrelated to TFA investigations, and they are frequently structured in formats that are inaccessible to non-technical users. To support advocates in safely using data exports for TFA-related cases, we built a filtering pipeline to exclude sensitive content such as media, certain settings, and background activity logs. However, this pipeline is currently platform-specific and brittle, as data export structures vary widely and frequently change, as discussed in Section 3.4. Future work could explore developing a client-side tool that helps survivors parse data exports locally, select relevant subsets of data, and share them with advocates via time-limited, end-to-end encrypted links. Such tools could automate redaction and support survivor agency in the data sharing process.

Abuse Potential of Data Exports. Safety risks to survivors during the request and handling of data exports are significant and must not be understated. Prior research highlights vulnerabilities in right of access processes, including weak authentication and susceptibility to social engineering attacks [10, 54, 55]. A major concern is that data exports are typically delivered to the primary account email, which may be compromised or shared with the abuser, potentially exposing sensitive data. Consultants must be trained before recommending data exports as a solution, and they should clearly communicate these risks during and after consultations. Safety planning should include secure storage, deletion, and a trauma-informed approach [17] to protect both the survivor and the consultant from harm.

Usability & Autonomy. Users often struggle to interpret their data exports [9, 82], and even our technical team found the data challenging to analyze (Section 3.4 & 4). Ambiguous terms like `"Summary of Device Events"` are rarely defined, and users may struggle to determine whether observed anomalies signal compromise has occurred. For TFA survivors, relying on experts to interpret this data reduces their autonomy and may introduce safety risks. Services could offer clearer exports through better documentation, in-line definitions, and customizable views to help users understand their data independently.

Use of Data Exports in Legal Contexts. Data exports offer a convenient way to access bulk account data and provide richer context for TFA than what is typically available through app interfaces. As such, data exports have the potential to serve as powerful evidence for survivors seeking legal remedies. However, data exports as they stand currently often include inconsistent and ambiguous data attributes that require subjective interpretation. Platforms should publish formal documentation of their data formats and clearly explain key attributes to reduce this burden. Finally, raising awareness of TFA and the evidentiary value of data exports among legal professionals is critical to ensuring better support for survivors.

8 Conclusion

Our study is the first to identify data exports as a tool for investigating account compromise within the context of interpersonal technology-facilitated abuse. We simulated four types of account-based attacks across six popular online services using researcher-controlled accounts, then analyzed the resulting data exports using a unified data model. We used qualitative content analysis to identify and characterize security-related information, and compared the depth and range of this data against what is present in account security interfaces (ASIs). By querying and inspecting data exports, we located traces of many, but not all, of our simulated attack steps across services. In contrast, ASIs frequently omit historical logins, changes to security settings, or device-level attribution. Our findings suggest that data exports can support more complete investigation of account compromise than is possible through ASIs alone. Future work should strive to create trauma-informed and privacy-preserving protocols to operationalize data exports within the clinical computer security setting.

Acknowledgment

We are deeply grateful to Sophie Stephenson and Michelle Jensen for their feedback on early drafts of this manuscript. We sincerely thank the anonymous reviewers for their insightful comments. We acknowledge the generous funding support from NSF award #2339679 and the Baldwin Wisconsin Idea Endowment. We acknowledge the Ho-Chunk Nation on whose ancestral lands we are grateful to work and live as guests. We deeply respect the knowledge embedded in the Ho-Chunk's custodianship of Teejop (DeJope) and recognize their continuing connection to land, water, and community here at the University of Wisconsin–Madison.

References

[1] Shiza Ali, Afsaneh Razi, Seunghyun Kim, Ashwaq Alsoubai, Joshua Gracie, Munmun De Choudhury, Pamela J. Wisniewski, and Gianluca Stringhini. 2022. Understanding the Digital Lives of Youth: Analyzing Media Shared within Safe Versus Unsafe Private Conversations on Instagram. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–14. doi:10.1145/3491102.3501969

[2] Shiza Ali, Afsaneh Razi, Seunghyun Kim, Ashwaq Alsoubai, Chen Ling, Munmun De Choudhury, Pamela J. Wisniewski, and Gianluca Stringhini. 2023. Getting Meta: A Multimodal Approach for Detecting Unsafe Conversations within Instagram Direct Messages of Youth. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1 (April 2023), 132:1–132:30. doi:10.1145/3579608

[3] Fatemeh Alizadeh, Timo Jakobi, Jens Boldt, and Gunnar Stevens. 2019. GDPR-Reality Check on the Right to Access Data: Claiming and Investigating Personally Identifiable Data from Companies. In *Proceedings of Mensch Und Computer 2019*. ACM, Hamburg Germany, 811–814. doi:10.1145/3340764.3344913

[4] Majed Almansoori, Andrea Gallardo, Julio Poveda, Adil Ahmed, and Rahul Chatterjee. 2022. A Global Survey of Android Dual-Use Applications Used in Intimate Partner Surveillance. *Proceedings on Privacy Enhancing Technologies* 2022, 4 (Oct. 2022), 120–139. doi:10.56553/popets-2022-0102

[5] Majed Almansoori, Mazharul Islam, Saptarshi Ghosh, Mainack Mondal, and Rahul Chatterjee. 2024. The Web of Abuse: A Comprehensive Analysis of Online Resource in the Context of Technology-Enabled Intimate Partner Surveillance.

In *2024 IEEE 9th European Symposium on Security and Privacy (EuroS&P)*. IEEE, Vienna, Austria, 773–789. doi:10.1109/EuroSP60621.2024.00048

[6] Jef Ausloos and Michael Veale. 2020. Researching with Data Rights. *Technology and Regulation* 2020 (2020), 136–157. doi:10.26116/techreg.2020.010

[7] Rosanna Bellini. 2023. Paying the Price: When Intimate Partners Use Technology for Financial Harm. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 872, 17 pages. doi:10.1145/3544548.3581101

[8] Rosanna Frances Bellini. 2024. Abusive Partner Perspectives on Technology Abuse: Implications for Community-based Violence Prevention. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1 (April 2024), 15:1–15:25. doi:10.1145/3637292

[9] Arthur Borem, Elleen Pan, Olufunmilola Obielodan, Aurelie Roubinowitz, Luca Dovichi, Michelle L. Mazurek, and Blase Ur. 2024. Data Subjects' Reactions to Exercising Their Right of Access. In *33rd USENIX Security Symposium (USENIX Security 24)*. USENIX Association, Philadelphia, PA, USA, 2865–2882.

[10] Luca Bufalieri, Massimo La Morgia, Alessandro Mei, and Julinda Stefa. 2020. GDPR: When the Right to Access Personal Data Becomes a Threat. In *2020 IEEE International Conference on Web Services (ICWS)*. IEEE, Beijing, China, 75–83. arXiv:2005.01868 doi:10.1109/ICWS49710.2020.00017

[11] California Consumer Privacy Act 2018. California Consumer Privacy Act of 2018. California Civil Code § 1798.100–1798.199. https://cppa.ca.gov/regulations/pdf/ccpa_statute.pdf

[12] California Privacy Rights Act 2020. California Privacy Rights Act of 2020. California Civil Code § 1798.100 et seq. https://cppa.ca.gov/regulations/pdf/prop24_text.pdf

[13] Rose Ceccio, Naman Gupta, Majed Almansoori, and Rahul Chatterjee. 2023. Analyzing the Patterns and Behavior of Users When Detecting and Preventing Tech-enabled Stalking. In *Proceedings 2023 Symposium on Usable Security*. Internet Society, San Diego, CA, USA. doi:10.14722/usec.2023.238140

[14] Rose Ceccio, Sophie Stephenson, Varun Chadha, Danny Yuxing Huang, and Rahul Chatterjee. 2023. Sneaky Spy Devices and Defective Detectors: The Ecosystem of Intimate Partner Surveillance with Covert Devices. In *Proceedings of the 32nd USENIX Conference on Security Symposium (SEC '23)*. USENIX Association, USA, 123–140.

[15] CETA. 2017. Clinic to End Tech Abuse (CETA) | Cornell Tech. https://www.ceta.tech.cornell.edu.

[16] Rahul Chatterjee, Periwinkle Doerfler, Hadas Orgad, Sam Havron, Jackeline Palmer, Diana Freed, Karen Levy, Nicola Dell, Damon McCoy, and Thomas Ristenpart. 2018. The Spyware Used in Intimate Partner Violence. In *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, San Francisco, CA, USA, 441–458. doi:10.1109/SP.2018.00061

[17] Janet X. Chen, Allison McDonald, Yixin Zou, Emily Tseng, Kevin A Roundy, Acar Tamersoy, Florian Schaub, Thomas Ristenpart, and Nicola Dell. 2022. Trauma-Informed Computing: Towards Safer Technology Experiences for All. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–20. doi:10.1145/3491102.3517475

[18] Dana Cuomo, Nicola Dell, Lana Ramjit, and Thomas Ristenpart. 2025. The Technology Abuse Clinic Toolkit. https://www.techabuseclinics.org. Accessed: 2025-04-12.

[19] Dana Cuomo and Natalie Dolci. 2021. New Tools, Old Abuse: Technology-Enabled Coercive Control (TECC). *Geoforum* 126 (2021), 224–232. doi:10.1016/j.geoforum.2021.08.002

[20] Dana Cuomo and Natalie Dolci. 2022. The TECC Clinic: An Innovative Resource for Mitigating Technology-Enabled Coercive Control. *Women's Studies International Forum* 92 (2022), 102596. doi:10.1016/j.wsif.2022.102596

[21] Alaa Daffalla, Marina Bohuk, Nicola Dell, Rosanna Bellini, and Thomas Ristenpart. 2023. Account Security Interfaces: Important, Unintuitive, and Untrustworthy. In *Proceedings of the 32nd USENIX Conference on Security Symposium* (Anaheim, CA, USA) *(SEC '23)*. USENIX Association, USA, Article 202, 18 pages.

[22] Discord. 2025. About Discord | Our Mission and Story. https://discord.com/company Accessed: 2025-04-12.

[23] Discord, Inc. 2025. Discord - Talk, Play, Hang Out. Mobile App. https://www.discord.com/ Accessed: 2025-04-10.

[24] Molly Dragiewicz, Jean Burgess, Ariadna Matamoros-Fernández, Michael Salter, Nicolas P. Suzor, Delanie Woodlock, and Bridget Harris. 2018. Technology Facilitated Coercive Control: Domestic Violence and the Competing Roles of Digital Media Platforms. *Feminist Media Studies* 18, 4 (2018), 609–625. doi:10.1080/14680777.2018.1447341

[25] Simon Ebbers, Fabian Ising, Christoph Saatjohann, and Sebastian Schinzel. 2021. Grand Theft App: Digital Forensics of Vehicle Assistant Apps. In *Proceedings of the 16th International Conference on Availability, Reliability and Security (ARES '21)*. Association for Computing Machinery, New York, NY, USA, 1–6. doi:10.1145/3465481.3465754

[26] Satu Elo and Helvi Kyngäs. 2008. The qualitative content analysis process. *Journal of advanced nursing* 62, 1 (2008), 107–115.

[27] European Parliament. 2016. General Data Protection Regulation. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AL%3A2016%3A119%3ATOC

[28] Facebook. 2025. Download Your Information. https://facebook.com/dyi.

[29] Jane Forman and Laura Damschroder. 2007. Qualitative content analysis. In *Empirical methods for bioethics: A primer*. Emerald Group Publishing Limited, 39–62.

[30] Diana Freed, Natalie N. Bazarova, Sunny Consolvo, Eunice J Han, Patrick Gage Kelley, Kurt Thomas, and Dan Cosley. 2023. Understanding Digital-Safety Experiences of Youth in the U.S.. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–15. doi:10.1145/3544548.3581128

[31] Diana Freed, Sam Havron, Emily Tseng, Andrea Gallardo, Rahul Chatterjee, Thomas Ristenpart, and Nicola Dell. 2019. "Is My Phone Hacked?" Analyzing Clinical Computer Security Interventions with Survivors of Intimate Partner Violence. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW (Nov. 2019), 202:1–202:24. doi:10.1145/3359304

[32] Diana Freed, Jackeline Palmer, Diana Minchala, Karen Levy, Thomas Ristenpart, and Nicola Dell. 2018. "A Stalker's Paradise": How Intimate Partner Abusers Exploit Technology. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–13. doi:10.1145/3173574.3174241

[33] Diana Freed, Jackeline Palmer, Diana Elizabeth Minchala, Karen Levy, Thomas Ristenpart, and Nicola Dell. 2017. Digital Technologies and Intimate Partner Violence: A Qualitative Analysis with Multiple Stakeholders. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 46 (Dec. 2017), 22 pages. doi:10.1145/3134681

[34] Andrea Gallardo, Hanseul Kim, Tianying Li, Lujo Bauer, and Lorrie Cranor. 2022. Detecting iPhone Security Compromise in Simulated Stalking Scenarios: Strategies and Obstacles. In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)* (Boston, MA, USA) *(SOUPS'22)*. USENIX Association, USA, Article 16, 22 pages. https://www.usenix.org/conference/soups2022/presentation/gallardo

[35] Alejandra Gómez Ortega, Jacky Bourgeois, and Gerd Kortuem. 2024. Sensitive Data Donation: A Feminist Reframing of Data Practices for Intimate Research Contexts. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference (DIS '24)*. Association for Computing Machinery, New York, NY, USA, 2420–2434. doi:10.1145/3643834.3661524

[36] Naman Gupta, Sanchari Das, Kate Walsh, and Rahul Chatterjee. 2024. A Critical Analysis of the Prevalence of Technology-Facilitated Abuse in US College Students. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3613905.3652036

[37] Naman Gupta, Kate Walsh, Sanchari Das, and Rahul Chatterjee. 2024. "I Really Just Leaned on My Community for Support": Barriers, Challenges, and Coping Mechanisms Used by Survivors of {Technology-Facilitated} Abuse to Seek Social Support. In *33rd USENIX Security Symposium (USENIX Security 24)*. USENIX Association, Philadelphia, PA, USA, 4981–4998.

[38] Sam Havron, Diana Freed, Rahul Chatterjee, Damon McCoy, Nicola Dell, and Thomas Ristenpart. 2019. Clinical computer security for victims of intimate partner violence. In *Proceedings of the 28th USENIX Conference on Security Symposium* (Santa Clara, CA, USA) *(SEC'19)*. USENIX Association, USA, 105–122. doi:10.5555/3361338.3361347

[39] Apple Inc. 2025. Apple Account. https://account.apple.com/

[40] Indian Ministry of Law and Justice. 2023. Digital Personal Data Protection Act. https://ethics.ncdirindia.org/asset/pdf/Digital_Personal_Data_Protection_Act_2023.pdf

[41] Instagram, Inc. 2018. Instagram for iOS. Mobile App. https://www.instagram.com/ Accessed: 2025-04-10.

[42] Christelle Kamaliza and Margaret Honda. 2020. *The State of Data Rights*. Technical Report. IAPP.

[43] Clare-Marie Karat, Robert Campbell, and Tarra Fiegel. 1992. Comparison of empirical testing and walkthrough methods in user interface evaluation. In *Proceedings of the 10th ACM SIGCHI Conference on Human Factors in Computing Systems*. ACM, 397–404. doi:10.1145/142750.142873

[44] Klaus Krippendorff. 2019. *Content Analysis: An Introduction to Its Methodology*. SAGE Publications, Inc., 2455 Teller Road, Thousand Oaks California 91320. doi:10.4135/9781071878781

[45] Jacob Leon Kröger, Jens Lindemann, and Dominik Herrmann. 2020. How Do App Vendors Respond to Subject Access Requests? A Longitudinal Privacy Study on iOS and Android Apps. In *Proceedings of the 15th International Conference on Availability, Reliability and Security (ARES '20)*. Association for Computing Machinery, New York, NY, USA, 1–10. doi:10.1145/3407023.3407057

[46] Niels Leenheer. 2024. User-Agent Reduction. https://nielsleenheer.com/articles/2024/user-agent-reduction/.

[47] Nicola Leschke, Daniela Pöhn, and Frank Pallas. 2024. How to Drill into Silos: Creating a Free-to-Use Dataset of Data Subject Access Packages. In *Privacy Technologies and Policy*, Meiko Jensen, Cédric Lauradoux, and Kai Rannenberg (Eds.). Vol. 14831. Springer Nature Switzerland, Cham, 132–155. doi:10.1007/978-3-031-68024-3_7

[48] Karen E C Levy. 2015. Intimate Surveillance. *Idaho Law Review* 51, 3 (2015), 679–693.

[49] Clayton Lewis, Peter G. Polson, Cathleen Wharton, and John Rieman. 1990. Testing a walkthrough methodology for theory-based design of walk-up-and-use interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Seattle, Washington, USA) *(CHI '90)*. Association for Computing Machinery, New York, NY, USA, 235–242. doi:10.1145/97243.97279

[50] Google LLC. 2025. Gmail - Email by Google. Mobile App. https://www.mail.google.com/ Accessed: 2025-04-10.

[51] Google LLC. 2025. Google Maps. Mobile App. https://www.maps.google.com/ Accessed: 2025-04-10.

[52] Alexander Löbel, René Schäfer, Hanna Püschel, Esra Güney, and Ulrike Meyer. 2024. Access Your Data... If You Can: An Analysis of Dark Patterns Against the Right of Access on Popular Websites. In *Privacy Technologies and Policy*, Meiko Jensen, Cédric Lauradoux, and Kai Rannenberg (Eds.). Springer Nature Switzerland, Cham, 23–47. doi:10.1007/978-3-031-68024-3_2

[53] Luis Marenco, Nicholas Tosches, Chiquito Crasto, Gordon Shepherd, Perry L. Miller, and Prakash M. Nadkarni. 2003. Achieving Evolvable Web-Database Bioscience Applications Using the EAV/CR Framework: Recent Advances. *Journal of the American Medical Informatics Association* 10, 5 (Sept. 2003), 444–453. doi:10.1197/jamia.M1303

[54] Mariano Di Martino, Isaac Meers, Peter Quax, Ken Andries, and Wim Lamotte. 2022. Revisiting Identification Issues in GDPR 'Right Of Access' Policies: A Technical and Longitudinal Analysis. *Proceedings on Privacy Enhancing Technologies* 2022, 2 (2022), 95–113. doi:10.2478/popets-2022-0037

[55] Mariano Di Martino, Pieter Robyns, Winnie Weyts, Peter Quax, Wim Lamotte, and Ken Andries. 2019. Personal Information Leakage by Abusing the GDPR 'Right of Access'. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. USENIX Association, USA, 371–385.

[56] Tara Matthews, Kathleen O'Leary, Anna Turner, Manya Sleeper, Jill Palzkill Woelfer, Martin Shelton, Cori Manthorne, Elizabeth F. Churchill, and Sunny Consolvo. 2017. Stories from Survivors: Privacy & Security Practices When Coping with Intimate Partner Abuse. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 2189–2201. doi:10.1145/3025453.3025875

[57] Meta Platforms, Inc. 2023. Facebook for iOS. Mobile App. https://www.facebook.com/ Accessed: 2025-04-10.

[58] Meta Platforms, Inc. 2025. Accounts Center. https://accountscenter.meta.com/

[59] Elizabeth A. Mumford, Poulami Maitra, Jackie Sheridan, Emily F. Rothman, Erica Olsen, and Elaina Roberts. 2023. Technology-Facilitated Abuse of Young Adults in the United States: A Latent Class Analysis. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 17, 3 (June 2023). doi:10.5817/CP2023-3-7

[60] National Congress of Brazil. 2018. Lei Geral de Proteção de Dados Pessoais. https://www.gov.br/anpd/pt-br/documentos-e-publicacoes/documentos-de-publicacoes/lgpd-en-lei-no-13-709-capa.pdf

[61] Jeremiah Onaolapo, Nektarios Leontiadis, Despoina Magka, and Gianluca Stringhini. 2021. {SocialHEISTing}: Understanding Stolen Facebook Accounts. In *30th USENIX Security Symposium (USENIX Security 21)*. USENIX Association, Berkeley, CA, 4115–4132. https://www.usenix.org/conference/usenixsecurity21/presentation/onaolapo

[62] Daniela Pöhn, Niklas Mörsdorf, and Wolfgang Hommel. 2023. Needle in the Haystack: Analyzing the Right of Access According to GDPR Article 15 Five Years after the Implementation. In *Proceedings of the 18th International Conference on Availability, Reliability and Security (ARES '23)*. Association for Computing Machinery, New York, NY, USA, 1–10. doi:10.1145/3600160.3605064

[63] Afsaneh Razi, Ashwaq Alsoubai, Seunghyun Kim, Shiza Ali, Gianluca Stringhini, Munmun De Choudhury, and Pamela J. Wisniewski. 2023. Sliding into My DMs: Detecting Uncomfortable or Unsafe Sexual Risk Experiences within Instagram Direct Messages Grounded in the Perspective of Youth. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1 (April 2023), 89:1–89:29. doi:10.1145/3579522

[64] Afsaneh Razi, Ashwaq Alsoubai, Seunghyun Kim, Nurun Naher, Shiza Ali, Gianluca Stringhini, Munmun De Choudhury, and Pamela J. Wisniewski. 2022. Instagram Data Donation: A Case Study on Collecting Ecologically Valid Social Media Data for the Purpose of Adolescent Online Risk Detection. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. ACM, New Orleans LA USA, 1–9. doi:10.1145/3491101.3503569

[65] Elissa M. Redmiles. 2019. "Should I Worry?" A Cross-Cultural Examination of Account Security Incident Response. In *2019 IEEE Symposium on Security and Privacy (SP)*. IEEE Computer Society, San Francisco, CA, USA, 920–934. doi:10.1109/SP.2019.00059

[66] Elissa M. Redmiles, Jessica Bodford, and Lindsay Blackwell. 2019. "I Just Want to Feel Safe": A Diary Study of Safety Perceptions on Social Media. *Proceedings of the International AAAI Conference on Web and Social Media* 13 (July 2019), 405–416. doi:10.1609/icwsm.v13i01.3356

[67] K. Andrew R. Richards and Michael A. Hemphill. 2018. A Practical Guide to Collaborative Qualitative Data Analysis. *Journal of Teaching in Physical Education* 37, 2 (April 2018), 225–231. doi:10.1123/jtpe.2017-0084

[68] Johnny Saldaña. 2013. *The Coding Manual for Qualitative Researchers* (2. ed ed.). SAGE Publ, Los Angeles, Calif.

[69] SAMHSA. 2014. SAMHSA's Concept of Trauma and Guidance for a Trauma-Informed Approach. https://www.health.ny.gov/health_care/medicaid/program/medicaid_health_homes/docs/samhsa_trauma_concept_paper.pdf

[70] Marija Schufrin, Steven Lamarr Reynolds, Arjan Kuijper, and Jorn Kohlhammer. 2021. A Visualization Interface to Improve the Transparency of Collected Personal Data on the Internet. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (Feb. 2021), 1840–1849. doi:10.1109/TVCG.2020.3028946

[71] Ali Fuad Selvi. 2019. Qualitative content analysis. In *The Routledge handbook of research methods in applied linguistics*. Routledge, 440–452.

[72] Julia Slupska and Leonie Maria Tanczer. 2021. Threat Modeling Intimate Partner Violence: Tech Abuse as a Cybersecurity Challenge in the Internet of Things. In *The Emerald International Handbook of Technology-Facilitated Violence and Abuse*, Jane Bailey, Asher Flynn, and Nicola Henry (Eds.). Emerald Publishing Limited, 663–688.

[73] Snap, Inc. 2025. Snapchat for iOS. Mobile App. https://www.snapchat.com/ Accessed: 2025-04-10.

[74] Sophie Stephenson, Majed Almansoori, Pardis Emami-Naeini, and Rahul Chatterjee. 2023. "It's the equivalent of feeling like you're in jail": lessons from firsthand and secondhand accounts of iot-enabled intimate partner abuse. In *Proceedings of the 32nd USENIX Conference on Security Symposium* (Anaheim, CA, USA) *(SEC '23)*. USENIX Association, USA, Article 7, 18 pages.

[75] Sophie Stephenson, Majed Almansoori, Pardis Emami-Naeini, Danny Yuxing Huang, and Rahul Chatterjee. 2023. Abuse Vectors: A Framework for Conceptualizing IoT-enabled Interpersonal Abuse. In *Proceedings of the 32nd USENIX Conference on Security Symposium (SEC '23)*. USENIX Association, USA, 69–86. https://www.usenix.org/system/files/usenixsecurity23-stephenson-vectors.pdf

[76] Madison Tech Clinic. 2021. Madison Tech Clinic. https://techclinic.cs.wisc.edu/.

[77] Kurt Thomas, Devdatta Akhawe, Michael Bailey, Dan Boneh, Elie Bursztein, Sunny Consolvo, Nicola Dell, Zakir Durumeric, Patrick Gage Kelley, Deepak Kumar, Damon McCoy, Sarah Meiklejohn, Thomas Ristenpart, and Gianluca Stringhini. 2021. SoK: Hate, Harassment, and the Changing Landscape of Online Abuse. In *2021 IEEE Symposium on Security and Privacy (SP)*. IEEE Computer Society, USA, 247–267. doi:10.1109/SP40001.2021.00028

[78] Clinic to End Tech Abuse. 2025. CETA | Resources. https://ceta.tech.cornell.edu/resources.

[79] Emily Tseng, Rosanna Bellini, Nora McDonald, Matan Danos, Rachel Greenstadt, Damon McCoy, Nicola Dell, and Thomas Ristenpart. 2020. The Tools and Tactics Used in Intimate Partner Surveillance: An Analysis of Online Infidelity Forums. In *29th USENIX Security Symposium (USENIX Security 20)*. USENIX Association, USA, 1893–1909.

[80] Emily Tseng, Mehrnaz Sabet, Rosanna Bellini, Harkiran Kaur Sodhi, Thomas Ristenpart, and Nicola Dell. 2022. Care Infrastructures for Digital Security in Intimate Partner Violence. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–20. doi:10.1145/3491102.3502038

[81] Tobias Urban, Dennis Tatang, Martin Degeling, Thorsten Holz, and Norbert Pohlmann. 2019. A Study on Subject Data Access in Online Advertising After the GDPR. In *Data Privacy Management, Cryptocurrencies and Blockchain Technology*, Cristina Pérez-Solà, Guillermo Navarro-Arribas, Alex Biryukov, and Joaquin Garcia-Alfaro (Eds.). Springer International Publishing, Cham, 61–79. doi:10.1007/978-3-030-31500-9_5

[82] Sophie Veys, Daniel Serrano, Madison Stamos, Margot Herman, Nathan Reitinger, Michelle L. Mazurek, and Blase Ur. 2021. Pursuing Usable and Useful Data Downloads Under {GDPR/CCPA} Access Rights via {Co-Design}. In *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*. USENIX Association, USA, 217–242.

[83] Miranda Wei, Eric Zeng, Tadayoshi Kohno, and Franziska Roesner. 2022. Anti-Privacy and Anti-Security Advice on TikTok: Case Studies of Technology-Enabled Surveillance and Control in Intimate Partner and Parent-Child Relationships. In *Proceedings of the Eighteenth USENIX Conference on Usable Privacy and Security*. USENIX Association, USA, 447–462.

[84] Adam Wolf. 2025. adamawolf/apple_mobile_device_types.txt. https://gist.github.com/adamawolf/3048717.

[85] Janis Wong and Tristan Henderson. 2018. How Portable Is Portable? Exercising the GDPR's Right to Data Portability. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers (UbiComp '18)*. Association for Computing Machinery, New York, NY, USA, 911–920. doi:10.1145/3267305.3274152

[86] Delanie Woodlock, Mandy McKenzie, Deborah Western, and Bridget Harris. 2020. Technology as a Weapon in Domestic Violence: Responding to Digital Coercive Control. *Australian Social Work* 73 (2020), 368–380. Issue 3. doi:10.1080/0312407X.2019.1607510

Appendix

We provide the complete codebook we developed as part of our qualitative content analysis of all the elements in six services in Fig. 8 below. These labels helped us characterize account security-related information (SRI) in data exports in Section 4. The codebook was developed with data from D_{Jan}, but also applied to D_{Feb}. We discuss generalizability in Section 3.4.

Activity Type Codes			
Level 1 Code	**Level 1 Definition**	**Level 2 Code**	**Level 2 Definition**
Authentication	Records about authentication including logins, logouts, sessions, trusted devices, and multi-factor authentication (MFA).	Login	Login instances, both successful and unsuccessful. Includes session initiation.
		Logout	Logout instances or session terminations.
		MFA	Multi-factor authentication (MFA) requests or challenges.
Settings	Records about account settings status or changes to account settings.	Email	Email addresses associated with an account or changes to emails. Includes verification during email change process.
		Password	Password metadata or changes to an account password.
		Profile	User profile settings (e.g., biographical information, profile picture, visibility) or changes to them.
		MFA	MFA status, or addition/modification/deletion of MFA methods. Includes verification during modification.
		Consent	Records about user consent to cookies, privacy policies, ToS, and EULAs.
Interaction	Records about users interacting with other users within the application, or creating user-facing content.	Chat	Records of a user sending messages, and posting text, media, or comments. Includes deleting or unsending any of the above.
		Emotion	Records of a user liking, saving, bookmarking, reporting, or blocking certain media content.
		Location Share	Records of a user enabling or disabling live location sharing (or receiving shared locations) with one or more contacts.
		Friend Modification	Friend/contact lists or records of a user adding, removing, or blocking a contact.
View	Record about users interacting with the application interface including opening the app, clicking on pages or content, or searching.	Search	Use of in-app search functions.
		View pages	Records of a user viewing or clicking on pages or content.
		App open	Records of a user opening or accessing the application, different from authentication as the user may be pre-authenticated when they access or open an application.
		Data Export	Data export requests or receipt.
Notification	Records about notifications sent to the user via push, email, or text; or user interactions with these notifications.		
Background Activity	Scheduled jobs, analytics and advertising, or other activity not driven by user actions.		
Other	Any remaining records of unknown or indecipherable meaning.		

Temporal Codes	
Event	Log entries corresponding to some historical, individual action or event, such as API calls or logins.
State	Metadata about ahistoric or persistent properties of an account or interface, such as biographical information or a list of enabled/disabled settings.

Figure 8: Codebook from our qualitative content analysis in Section 4. Note that each set of Level 2 codes had an additional "Other" category for data elements that did not clearly adhere to our codes, but were not prevalent enough to justify an additional category.