Data Stewardship in Clinical Computer Security: Balancing Benefit and Burden in Participatory Systems

EMILY TSENG, Cornell University, USA

ROSANNA BELLINI, Cornell University, USA

YEUK-YU LEE, Cornell University, USA

ALANA RAMJIT, Cornell University, USA

THOMAS RISTENPART, Cornell Tech, USA

NICOLA DELL, Cornell Tech, USA

The mass collection and reuse of social data requires a reimagining of privacy and consent, with particular attention to the (in)equitable distribution of benefits and burdens between researchers and subjects. Instrumenting frontline clinical services to collect and steward data might mitigate the exploitation inherent to data collection—with attention to how subjects can meaningfully participate in stewardship. We explore participatory data stewardship in the context of clinical computer security for survivors of intimate partner violence (IPV). Via semi-structured interviews with IPV support workers, we explore how data are produced within the IPV care ecosystem at the Clinic to End Tech Abuse (CETA). We then conduct design provocations with clients of IPV services and their support workers, exploring possibilities for participatory data mechanisms like open records and dynamic consent. We find participation in data stewardship may benefit clients through improved agency, self-reflection, and control of self-narrative, and that incurred burdens may be alleviated by enlisting trusted stewards. We close with future work for CSCW interrogating how knowledge of digital-safety harms can and should be produced from clinical encounters, towards more equitable ways of knowing.

CCS Concepts: • **Human-centered computing** → **Empirical studies in collaborative & social computing**.

Additional Key Words and Phrases: digital safety, intimate privacy, data stewardship, open science, participation

1 INTRODUCTION

First-hand perspectives from vulnerable communities are necessary for meaningful and positive social change—but gathering these perspectives risks exploiting these communities. This tension is exacerbated by recent evolutions in our knowledge infrastructures towards mass collection and reuse of social data: trends known as *big data* and *open science* [36]. These trends have enabled large-scale scientific advancements, but have also deepened existing power differentials, requiring innovative approaches to core tenets of research ethics like autonomy, privacy, and consent [56, 76].

Authors' addresses: Emily Tseng, Cornell University, New York, NY, USA; Rosanna Bellini, Cornell University, New York, NY, USA; Yeuk-Yu Lee, Cornell University, New York, NY, USA; Alana Ramjit, Cornell University, New York, NY, USA; Thomas Ristenpart, Cornell Tech, New York, NY, USA; Nicola Dell, Cornell Tech, New York, NY, USA.

39

One alternative approach to knowledge production is *participatory data stewardship* [1], in which communities are directly involved in the collection, analysis, and sharing of data. Community archives and citizen science have shown increasing participation in data stewardship can both restore agency to data subjects and enrich research findings [29, 53]. But these practices do not completely eliminate exploitation, as they can burden communities with the labor of collecting data, storing it, and making ethical judgments about its use and re-use. To truly foster justice, these systems must carefully distribute benefit and burden between researchers and subjects [32, 69].

Instrumenting clinical sites as research observatories is one tactic for ensuring that research encounters directly benefit subjects—by situating those encounters where hard-to-reach populations already seek care. This approach is gaining traction in health, where researchers use secondary analyses of clinical records to study population-scale mental and physical health [4, 11, 45, 57, 61, 85]. However, clinical observatories require difficult decisions about meaningful consent and appropriate re-uses of stewarded data [16]—an open space for advancement in participatory data stewardship.

This paper explores these tensions in the context of clinical computer security for survivors of intimate partner violence (IPV). IPV survivors face complex tech-mediated harms like intimate surveillance [60], harassment and doxxing [42], and financial abuse [9]. Prior work has introduced computer security clinics as a form of care infrastructure, where survivors receive expert security and privacy support [26, 41, 48, 88, 90]. In a two-part study situated within one such clinic, the Clinic to End Tech Abuse (CETA) in New York City, U.S., we extend this literature towards digital-safety research, by examining through co-design how the clinic's data handling mechanisms might give survivors more agency in how knowledge about their digital safety is produced and circulated.

Following a trauma-informed approach [23], we first mapped the current landscape of data collection, analysis, and sharing in the IPV support ecosystem, through semi-structured interviews with 28 support workers (4). This baseline understanding informed the design of our second phase: design provocations exploring possibilities for participatory data infrastructures in CETA (5). Through semi-structured design activities with 7 CETA consultants and 7 clients, we explored two key concepts in participatory data systems: the ability to view and edit one's own data (*open records*), and the ability to control who one's data are shared to (*dynamic consent*).

Our findings are summarized in Table 1. Within the current IPV service ecosystem, we found a series of important yet fragmented data flows. These clinical encounters are sites of highly personal disclosure from clients (and occasionally support workers), and the resulting records often do contain hard-to-reach knowledge of abuse experiences. To appropriately handle this information, support workers rely on informal practices and personal judgment calls, for which they crave improved frameworks and tools—especially for informed consent. Prompted to consider open records and dynamic consent, participants reflected that these mechanisms could afford much-appreciated transparency and control over their data. Still, they cautioned that these affordances might introduce too much extra burden, especially in the time and effort required to vet data requests from secondary researchers, and that there remained a role for a trusted data steward.

Based on these findings, we argue participatory data stewardship is worth pursuing, for the benefits to agency, self-reflection and control it confers to clients and consultants, and for its potential to improve the data available for wider research—including for helping service organizations attain funding. However, to adapt current frameworks of participation in data stewardship to digital-safety contexts like the clinic, we need further work interrogating two core tensions between the literature and our findings: how to handle conflicts in individual vs. collective gains in data stewardship (6.1.1); and how to balance burden and supervision between researchers and data subjects (6.1.2).

In addition to theoretical developments, we see pragmatic next steps improving the mechanisms of informed consent (6.2.1), and alleviating the burdens of participatory data stewardship through careful and human-centered design of computational tools for stewards (6.2.2). Finally, at a meta

Phase	Findings
Understanding the current landscape of data stewardship in frontline IPV services (4)	Clinical encounters provide rich, yet imperfect accounts of survivors' experiences (4.2.1)
	Support workers balance the need for data against the need to protect clients' privacy (4.2.2)
	Support workers also consider their own privacy in documenting clinical encounters (4.2.3)
	Support workers find current informed consent procedures rigid and perfunctory (4.2.4)
	Clients seek access to their data; but support workers caution against unfettered access (4.2.5)
Exploring mechanisms for participatory data stewardship (5)	Clients also find current informed consent frameworks rigid and perfunctory (5.2.1)
	Open records give survivors an avenue for self-reflection and control of self-narrative (5.2.2)
	Granular selective sharing in dynamic consent accommodates diverse preferences (5.2.3)
	With granular and self-directed information control, a role remains for trusted stewards (5.2.4)

Table 1. Summary of findings.

level, we urge CSCW towards acknowledgement of the emotion work required to conduct participatory design of participatory technologies in digital-safety research, and how we can sustain long-term research commitments in this area (6.2.3). In sum, we contribute:

(1) A co-design study with IPV survivors and their support workers examining concrete mechanisms for participation in stewarding knowledge of IPV.
(2) Theoretical extensions for how the literature in participatory data stewardship imagines individual vs collective benefit, and the dialectic between researcher and data subject.
(3) Practical future directions for CSCW and HCI, improving how participatory data stewardship can be applied to digital-safety research.

2 BACKGROUND AND RELATED WORK

Our work addresses CSCW's ongoing interest in our changing knowledge infrastructures [6, 14, 17, 54, 72], and analyzes how to adapt them towards research in a critical context: digital safety for survivors of intimate partner violence (IPV). Here, we first review the CSCW literature on one emerging knowledge practice, *data sharing and reuse*, and how it is increasingly incentivized across science—even as it poses ethical challenges for marginalized communities (2.1). We then describe an alternative approach, *participatory data stewardship*, which aims to give data subjects greater agency over the use of their data (2.2). We discuss the potential of participatory data stewardship in IPV, and motivate our study of two mechanisms: *open records* and *dynamic consent* (2.3).

2.1 Increasing imperatives towards data sharing and reuse

A growing arc of CSCW traces the evolution in our knowledge infrastructures towards big data and Open Science—and accompanying ethical questions in how we study and share knowledge of the world. In 2013, Edwards et al.[36] described a range of emerging knowledge practices—crowdsourcing, personal informatics, massive online open courses (MOOCs), etc.—made possible by wide adoption of personal computing, the Internet, and machine learning (ML). These practices signal infrastructural transitions in science: from knowledge as a fixed product to knowledge in perpetual motion; from individual expertise to the wisdom of crowds.

We focus on one influential practice: sharing and reusing data. Such activities are core to Open Science: a movement to "make research *and data* accessible to all" [65]. Driven by the values of democratization, transparency, and replicability, Open Science envisions research data and findings

as shared goods, and researchers as accountable not only to peers and funders, but also to the public. Towards this vision, institutions and funders have invested in making research data **F**indable, **A**ccessible, **I**nteroperable, and **R**e-usable, via the FAIR Data Principles [95]. Proponents argue FAIR accelerates science by fostering collaboration and streamlining research funding. Popular in ML, FAIR is gaining ground in qualitative research, where proponents argue data reuse enables productive recontextualization of social phenomena [64]. In fact, Edwards et al. 2013 urged the development of long-term, large-scale, and collaborative qualitative research infrastructure. [36]

Ethical challenges in sharing data. FAIR requires researchers to grapple with not only how to ethically *collect* data, but also how to ethically *share* it with secondary researchers. A scientific enterprise built on mass data circulation challenges values of autonomy, privacy, and trust [66, 76, 91]. Research ethics has long held that people must give voluntary and informed consent to research participation [56, 58]. But when data can be infinitely reused, people cannot know the future uses to which they are consenting. The principle of anonymity, too, is no longer enough. Scrubbing datasets of personally identifiable information was once thought sufficient for confidentiality, but mass open data has made it easy to infer sensitive attributes from widely available digital traces [8].

Open Science's challenges to consent and privacy belie a broader problem: a growing power imbalance between researcher and subject. Amassing data on society confers enormous power to those who store and analyze it—power which does not always reciprocally benefit data subjects, particularly when captured as capital. For example, pharmaceutical companies build intellectual property from genetic and biomedical data in government-sponsored biobanks, while the people who contribute their flesh and blood receive little to no renumeration [40, 79]. Fundamentally, to collect data is to categorize the world, and per Bowker and Star [15], the act of classification "*valorizes some point of view and silences another*". As datasets drive systems for automated social decisions, we risk concretizing the systemic oppression of marginalized people (cf. [38, 67, 80]).

Taking into account the imperatives of data reuse and sharing under Open Science, along with the ethical challenges posed by those same practices, we locate the present study on a critical frontier for CSCW: ethical reimagination of our evolving knowledge infrastructures. To emphasize the moral accountability and long-term commitment required of research today, we follow the Ada Lovelace Institute and reframe the activities of data collection and sharing as *data stewardship* [1].

2.2 Fostering meaningful participation in data stewardship

Towards redistributing the power imbalance implied in data sharing, scholars have proposed *participation* as a design goal for data stewardship. A participatory ethos envisions that with greater agency in how their stories are studied and shared, marginalized people may leverage science to improve their condition [69]. Scientific knowledge improves, too: as knowledge emerges from contextual relationships between researchers and the researched [19, 64], involving data subjects in data stewardship might enable greater insight, developed *with* and not *on* people and communities.

How precisely to involve data subjects in participatory data stewardship is the subject of a long literature in participatory design. At issue is when participation is *meaningful*: when it can help the powerless, and when it can instead be co-opted by the powerful; or when it can levy greater burdens on data subjects than benefits. [32, 33, 88, 89, 92] Asking people to make constant decisions about their data can give them greater control [91], and may increase their data literacy [55]. However, citizen science shows it takes work to ready community-collected data for reuse—and this is a lot to ask of people who are already marginalized and underserved [69]. One approach is to abstract the burden of decision-making to a collective; still, community governance poses its own legal, logistical, and technical problems [43]. In computing, where stewarded data rapidly
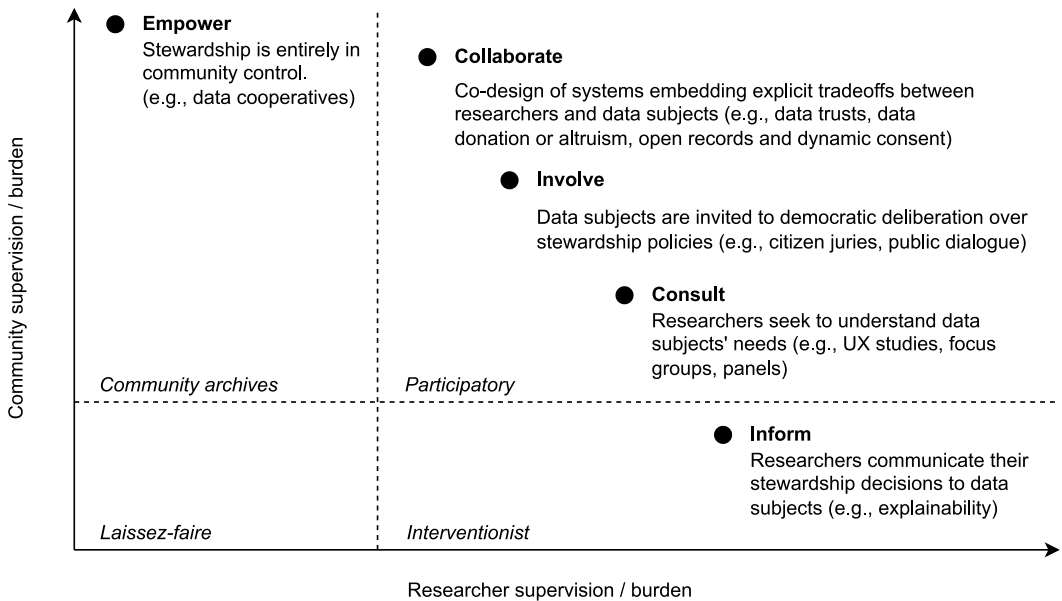
Fig. 1. Heuristic diagram of tradeoffs in participatory data stewardship in the current literature. Maps the Ada Lovelace Institute's typology of mechanisms for participation [1] (black dots) with Jo and Gebru [53]'s spectrum of community vs researcher supervision (axes). Both frameworks encourage moving up and to the left: towards enabling communities to own and operate archives of their experiences. Our study uses design provocations of mechanisms under *collaborate*, open records and dynamic consent, to explore participation in stewarding knowledge of digital-safety harms, such as those in intimate partner violence (IPV).

become deployed technologies, it is especially important to work towards meaningful participation for as many people as possible (cf. [10] on "participatory" artificial intelligence).

Synthesis of frameworks for participation in data stewardship. In this work, we consider two frameworks for data stewardship with complementary approaches to meaningful engagement: (1) Jo and Gebru's [53] spectrum of supervision in data collection; and (2) the Ada Lovelace Institute's [1] framework for "*participatory data stewardship*".

In the first framework, Jo and Gebru describe researcher supervision over data collection as a spectrum between two extremes: at one end, "*laissez-faire*" approaches like the unstructured web scrapes used in modern ML, and at the other, "*interventionist*" approaches like those of professional curators, who define and control what information is collected. As an alternative, Jo and Gebru describe "*community archives*", in which data subjects organized as collectives are deeply involved in efforts to record knowledge of their lives.[1] The second framework, from the Ada Lovelace Institute, more closely examines community-involved approaches to data stewardship. This framework extends Arnstein's ladder of citizen participation [5] to outline mechanisms for data governance along a spectrum of data subject or community control.

Our synthesized heuristic diagram (Figure 1) maps Lovelace's typology of mechanisms for participation on two axes: for researchers and communities each, the degree of supervision required and burden bestowed. Supervision and burden are coupled together along these axes to highlight

---

[1]Examples include activists documenting feminicides in Latin America [30], the Patient-Led Research Collaborative formed to understand long COVID, and LunaDNA, a genomics research cooperative.

how increased control of a project can also demand time, effort, and emotional labor (cf. Dourish et al. [33]). At one corner is the category *inform*: mechanisms where data subjects are told about the research activities conducted on them (e.g., artifacts for transparency, like datasheets). Here, researcher supervision is maximal and community supervision is minimal. At the other end is the category *empower*: mechanisms by which communities operate the archive and control the research conducted with it. Here, community supervision is maximal and researcher supervision is minimal. In between these extremes lie what we call *participatory* mechanisms, where the distribution between researcher and community supervision is more balanced. Here we place Lovelace's *consult*, *involve*, and *collaborate* types, in increasing order of community supervision.

With these tradeoffs in mind, how should designers of data stewardship systems choose an approach? The goal, both frameworks argue, is to be as close to community control of the system as possible: to move up and to the left on Figure 1. How precisely to get there is a matter of selecting complementary mechanisms: "*there is no single 'right' way to do participation*", the Lovelace report urges, and the typology is not mutually exclusive [1]. Instead, the best-fit combination of mechanisms should be developed through collaborative processes like co-design, through which researchers and communities can both gain insight on not only the immediate effects of a proposed mechanism, but also its wider ecosystem of construction, maintenance, and cultural significance [71, 78, 97]—including the burdens it may create for all parties. We take up this call, to examine through co-design how participatory data stewardship might improve knowledge of digital safety in IPV, and improve agency for IPV survivors.

2.3 Participatory data stewardship and digital-safety research in IPV

Our work explores the potential of participatory data stewardship for improving knowledge in a particularly challenging domain: digital safety in IPV. A widespread social ill [22], IPV is characterized as a pattern of "*coercive control*" [81, 82] in a current or former intimate partnership that manifests in many forms of harm, including physical, psychological, financial and sexual abuse.

Research on survivors' experiences is important for addressing their needs, and mitigating gender-based violence broadly. Already, research has helped establish legislative protections [84, 86], improved law enforcement responses [25, 70], enhanced economic support [13], and increased public awareness [21]. However, this research has required updating traditional methods, like surveys and interviews. Increasingly, digital technologies exacerbate IPV, giving an abuser new tools to intimidate, harm, stalk, and harass their victims [26, 34, 42, 47, 87]. As a result, survivors are often hard to reach: the measures they take to evade their abusers may involve rapidly changing their contact information, and avoiding unsolicited reach-outs like recruitment messages [88].

Instrumenting IPV support services as data stewards. One possible way forward is to study IPV by equipping frontline IPV support services to collect, analyze, and share data about their clients. Such an approach is increasingly used across health contexts. Clinical encounters—the dyadic interactions between people seeking care and their care providers—were previously seen as highly private arenas for personal disclosure. But with Open Science, secondary analyses of electronic medical records (EMRs) have become an important source of data on human well-being [11, 45, 57]. In mental health, analyses of therapist-patient exchanges on digital platforms have explored hard-to-reach topics like linguistic markers of suicidality [4, 31, 61] and psychological responses to COVID-19 [51]. This approach hopes that by meeting care seekers where they are, researchers can study rare or stigmatized conditions while minimizing the burden research imposes.

In theory, equipping IPV support services to steward data would offer a rich observatory on this underreported societal problem [28, 63], enabling research towards mitigation at the scale of what EMRs and digital care platforms have done for physical and mental health. Already researchers

have pursued alternative ways to measure IPV's impact, via estimation from medical records [75] or search engine queries [83]. If collected in meaningfully participatory ways, data from frontline services could join these efforts, to better inform mitigations in law, policy, and technology design.

However, the question of meaningful participation in stewarding IPV data is especially complex. While the literature emphasizes community control of data (Figure 1, top-left), to our knowledge, no prior work has explored balancing community and researcher supervision in data stewardship for IPV and digital safety. Research participation generally is known to offer benefits for survivors: interviews can be spaces for personal insight and healing [24, 44, 77], and survivors have reported enthusiasm for broader goals, like helping others spot signs of abuse [39] and improving advocacy services [93]. Still, the burdens of participation are equally many. Recounting abuse experiences may re-traumatize survivors [23], and as has been observed in mental health [16], survivors may have only momentary encounters with services, creating challenges for informed consent. At worst, survivors may feel coerced to consent to research to receive service. To find a way through these challenges, we study in this work how to balance these benefits and burdens, towards remaking a frontline support service for IPV survivors as a participatory data steward.

3 RESEARCH CONTEXT: THE CLINIC TO END TECH ABUSE (CETA)

We studied the complexities of data stewardship for IPV via embedded research with the Clinic to End Tech Abuse (CETA). Established in 2018, CETA provides services for IPV survivors experiencing tech-enabled abuse. CETA sits within a broader ecosystem of IPV support services in New York City, where professional advocates help survivors find everything from legal services to psychotherapy. These advocates (*IPV professionals* hereafter) refer survivors with tech-related problems to CETA.

CETA volunteers (*consultants*) are trained in computer security and trauma-informed care, and provide 1:1 support for IPV survivors (*clients*) [41, 48]. As of 2023, the clinic sees 4-6 clients per week for problems ranging from password compromise to smart home surveillance. For each client, all information pertaining to their case is stored in a record accessible only to the client's consultant team and the clinic's leadership team (see [90] for detail on CETA's data architecture). These records include an anonymized version of the referral form filled out by the client's IPV professional. Clients may also optionally provide demographic data on their age, race, sexuality, and gender. Within an encounter, consultants take notes on client concerns and suggested approaches. These notes are also stored in case folders, alongside appointment summaries.

Like many clinical sites, CETA conducts research alongside providing services. Consultants ask clients if they would like to participate in research by having their clinical encounter audio-recorded and transcribed. For clients who consent, case records—including transcripts—are anonymized and made available to the CETA research team. This data has enabled CETA to publish research on less-studied dimensions of IPV [9, 41, 48, 90]. Over the years, CETA has also received numerous requests for data from other researchers, lawmakers and advocacy groups.

CETA has typically responded to these requests by carefully vetting researchers before providing access to data, or by providing consultants' perspectives as experts in lieu of client records. However, advances in participatory data stewardship invite inquiry into how the clinic might enable participatory forms of research and knowledge dissemination. In this work, we use CETA as a case study to explore how digital-safety clinics like it can respond to requests for data in ways that enable participation for IPV survivors, while stewarding sensitive data in ethical ways. In this, we seek concrete lessons for CETA itself, for the broader ecosystem of frontline service and research in IPV, and for CSCW's interest in our changing knowledge infrastructures.

Open records and dynamic consent. Of the mechanisms for participatory data stewardship reviewed in Figure 1, we focus on two especially popular in clinical services analogous to CETA (cf.

[35, 91, 94]): *open records* and *dynamic consent*. The first, *open records*, refers to mechanisms that provide data subjects with access to records about them.[2] The second, *dynamic consent*, refers to mechanisms that enable people to opt in or out of data collection over the lifetime of their data.

Open records and dynamic consent are core building blocks of participation in data stewardship. These mechanisms are becoming more widely used across health, via biobanks and mobile health research tools like Apple's ResearchKit [35, 94]. They have also gained purchase in other contexts: the ability to view and take control of one's own digital traces is enshrined as a basic consumer protection in the European Union's General Data Protection Regulation. In social media, recent scholarship has proposed dynamic consent as a lens for safer online interaction [52], and as a tool for people to meaningfully participate in online experiments—or to refuse them [98]. Due to the growing popularity of these two mechanisms, our study uses them as prompts for exploring the design space of participatory data stewardship in IPV.

4 PHASE I: MAPPING CURRENT CHALLENGES

To begin our co-design process, we first explored through formative interviews the current landscape of data collection, analysis, and reuse in IPV services.

4.1 Methods

Following trauma-informed research practice [23, 42], we began by interviewing proxies for survivors prior to approaching survivors for direct work (Phase II, Section 5). We interviewed 19 volunteer CETA consultants and 9 IPV professionals who refer clients to CETA. Participants were asked to describe how they manage data about their clients, and how they handle requests for client data from external organizations. Interview protocols are provided in Supplementary.

Recruitment. We recruited participants who have firsthand experience working with IPV survivors, coordinating teams of frontline workers, or overseeing monitoring and evaluation for IPV services. To recruit CETA consultants, we reached out via email and Slack to 20 consultants randomly sampled from CETA's active roster. Upon consultation with CETA, we did not offer financial compensation to CETA volunteers. To recruit IPV professionals, we reached out to 10 advocates snowball sampled from CETA partners. We offered USD$25 giftcards to professionals as a token of appreciation, except for participants who were government workers prohibited from accepting gifts. In total, 19 consultants and 9 professionals agreed to participate in 60-minute Zoom interviews.

Analytical approach. Transcripts and notes from the 28 interviews were analyzed using a reflexive thematic analysis adapted from Braun & Clarke [18]. Through multiple rounds of iterative and collaborative coding, we first developed 40 codes from the consultant data (e.g., *data-driven performance evaluation*, *wall of legalese*, *consultants want control of their narrative*). Next, the first author coded the IPV professionals' data, generating a new codebook of 41 codes (e.g., *threat of subpoena*, *streamlining services*, *case reviews*). Related codes across the two codebooks were clustered and synthesized into Section 4.2. Codebooks are reported in Supplementary.

Ethics and researcher positionalities. This was a reflective and reflexive study drawing on the authors' collective depth of experience in the subject. Reflexive research has been used in CSCW and critical HCI to produce rich insights on complex and emotionally difficult phenomena like IPV [37, 41, 74]. All authors have extensive experience in researching digital safety and IPV, and act as volunteers in CETA as part of a larger 30+ person team. Some are represented in the data. Our

---

[2]Open records have recently been instilled in the U.S. healthcare system through the 21st Century Cures Act of 2021, which created a federal mandate for patients to have access to their medical records without delay. [27, 68]

analyses aim to critically appraise and improve CETA's services, towards improving digital safety and towards novel insights on these new forms of data infrastructure for the CSCW community.

We received IRB approval for all study procedures. Potential participants were assured that agreeing to or declining the study would not impact their relationship with CETA. We additionally ensured no consultant was interviewed by their supervisor. In the writing of this paper, we took precautions to ensure participant anonymity, e.g., removing potentially identifying phrases from quotes. Each author also had a self-care plan for mitigating vicarious trauma and burnout.

4.2 Findings

Our findings (Table 1) illuminate how frontline IPV services collect, analyze, and reuse case data. Here, we use CX for consultants and PX for professionals, where X is the participant number. Where we refer to one subgroup, we use *consultants* or *professionals*; for both, we use *support workers*.

4.2.1 Clinical documentation provides rich, yet imperfect accounts of survivors' experiences. From the moment a survivor contacts a support service, that service begins to make decisions about what information about them needs to be collected and shared. Across disparate groups, one common challenge is how to document a client's initial **disclosure** of the abuse they experienced.

Support workers in our study and in prior work [41, 88, 90] indicate the disclosure part of an encounter can be lengthy and emotionally taxing. Even when disclosures are not related to the services sought, support workers give survivors ample room to share their stories, to create trust and rapport, and reinforce what might be a quasi-therapeutic relationship for survivors.

Often, these highly personal disclosures contain vital information about the client's situation, used by support workers to tailor the rest of the encounter. But support workers often struggle to record these disclosures quickly and thoroughly enough. As one described:

> "*It's kind of hard when a client is spilling their soul out to you and then all of a sudden they hear typing, and you're like I'm so sorry, my brain is the size of a pea right now, I really need to write this down to remember all the things that you told me.*" (P08)

Much like doctors adopting shorthand to write patient notes, support workers use keywords or other schemes to quickly and accurately record what clients tell them, while still remaining present in the encounter. This record-taking often involves writing referral emails for the client, or searching for information online. All professionals described this balance—between thorough documentation, timely information searching, and staying attentive—as core to their work.

Given the difficulty of documenting clinical encounters 'in-the-moment', support workers also say it is hard to reconstruct these encounters from collected data, no matter the medium. Written notes might lack important information like cadence and tone; and unstructured observational data, like transcripts of audio recordings, might be insufficient for systematic analysis. The imperfect nature of the data affects not only practice but also efforts at research: one IPV professional said that to investigate specific constructs, they wanted quantitative surveys, "*to get out of a hole of diverse vocabulary, or even just typos*" (P09).

4.2.2 Support workers balance the need for data against the need for privacy. Data collection also challenges another core concern: protecting clients' privacy. Support workers make many highly subjective judgment calls about how much data to collect, while also contending with the reality that quality data collection can be critical to their organizations' funding. Within the IPV threat model (Section 3), abusers may relentlessly pursue survivors' digital footprints. Information disclosed and recorded in a clinical encounter may be a particularly enticing target. Despite computer security best practices, organizations cannot entirely eliminate the risk that recorded data ends up disclosed.

Support workers described a pastiche of formal and informal practices for minimizing the data collected. The aspirational standard is to simply not record more information than needed—but definitions of minimum viable data vary, and rely on support workers' "*in-the-moment*" decisions. One professional said that over the last few years, her organization had collected less and less information. Their case records presently serve mostly to document the services rendered to the client: "*It might be like, 'spoke about housing', a broad category like that. That's really it.*" (P02). Similarly, CETA consultants try to limit the data they collect to what they consider to be relevant to tech abuse, CETA's remit—but in practice, the line between what is and isn't relevant to tech abuse is often fuzzy. In making these judgment calls, participants say they consider what would happen if the data were leaked—especially to the abuser, via hostile subpoena by an abuser's legal team.

To complicate these judgment calls, participants are very aware that their organizations need quality data to improve and sustain their programs. One IPV professional's organization did not want to collect clients' addresses, but as an organization supporting a diversity of communities in New York City, they needed *some* demographic data to develop more equitable services, or apply for specific grants. Ultimately, the organization decided to record high-level data like ZIP codes, primary language spoken, and racial or ethnic background, and to use it for these purposes only.

Nonetheless, service delivery often *requires* collecting personal information, e.g., phone numbers used to contact survivors, or photo identification used for security purposes. Support workers try to be flexible in these requirements for very wary clients: some are trained to ask "*is this the name you would like us to use?*" instead of "*is this your name?*". Clients are then identified in databases by this pseudonym only: as one professional described, *"There's a lot of Beyoncés in our system"* (P02).

4.2.3 Support workers may also require privacy protections. Clinical encounters are dyadic exchanges between care seekers and caregivers. We found that records from these encounters may contain sensitive disclosures from not only clients, but also their support workers, who worry that their own disclosures could be exposed. In addition, support workers report concern that any data on their actions during client sessions might subject them to negative sanctions over work performance by a supervisor, or judgment from their colleagues.

Personal disclosures from support workers may occur naturally, as part of how they empathize with and emotionally support a client. A CETA consultant recalled one such incident:

> "*It got really emotional ... I wanted to tell her I understand how hard it is, and I can relate ... after that call, I remember feeling really bad that I told the client [about my experience of abuse].*" (C14)

Once recorded, support workers' personal stories or information can be leaked in the unlikely event of a data breach or a hostile subpoena. More subtly, records of support workers' actions—e.g., disclosures, questions asked and not asked, referrals made and not made—can also expose them to unfair judgments from their supervisors or colleagues. Recalling negative experiences with corporate jobs, several participants said they were wary that records of their actions could help create uncomfortably surveillant work environments where "*supervisors are constantly looking over your shoulder*" (C07). While CETA consultants know these concerns are mitigated by their status as volunteers, they are still concerned about reputational harms:

> "*I'd be afraid [a record of my mistake] would be used in training. Then [my colleagues] would be like, 'Oh, this guy. He really messed up.' They might not see what I've done to improve since.*" (C15)

Still, most participants see value in recording *some* data on their own roles in clinical encounters, for training and improvement—if accompanied by the enforcement of "*a culture where constructive, supportive critique is embraced*" (C07). For several professionals, recording sessions for training and supervision is already accepted practice. In fact, for these purposes, they actively seek more data with greater detail. Audio recordings, for example, contain important detail that transcripts lack:

"*tone and voice, and pauses, and certain qualitative aspects you can't really get otherwise*" (P07). We discuss these tensions further in Sections 6.1 and 6.2.2.

4.2.4 Support workers find current consent frameworks rigid and perfunctory. All organizations surveyed in this work require clients to give informed consent, to receive services and/or participate in research—and many support workers describe this is particularly challenging.

Our participants stressed that informed consent is particularly important for IPV survivors, who may be denied the ability to make decisions about their own life by an abuser [93]. Thus an important tenet of client-centered care in IPV is to properly inform clients, and give them control over their personal information, particularly when decision-making involves safety risks.

But when obtaining consent, support workers can struggle to balance informing clients against overwhelming them. As one professional says, they seek to give "*as much information as possible, without getting so into the weeds that they tune out and click away*" (P09). However, support workers often cannot gauge whether clients truly understand what they were agreeing to. Recalling CETA's consent process—a verbal consent, with a written form available on request—one consultant said:

> "*It's a lot of words, a lot of legalese... I've never had anyone ask a question about it. So I don't have any way to know that they have taken it in, thought about it, understood it.*" (C02)

To some support workers, the placement of the consent process—traditionally done at the start of an encounter—could inadvertently "*put some pressure [on clients] to go ahead and consent*" (C11), so they could skip quickly to receiving services. As one consultant described:

> "*It's like, "Oh, yeah. I consent, I consent, I consent. Let's get to the important stuff." No judgment, but it's like the important thing that they're there for is to get this help, and so the consent process is sort of an obstacle to that rather than an important part of it.*" (C17)

To these support workers, clients' ritualistic responses— "*check the box, sort of like that*" (C05)— should not be interpreted as disinterest in informed consent, but rather as an indication that clients pragmatically seek to focus on more pressing matters. We unpack this further in Section 6.2.1.

4.2.5 Clients seek their own data; but support workers caution discretion. Many consultants and professionals recalled instances where clients sought access to their own case records, including recordings. While there can be clear benefits to giving clients this information, such as the accumulation of evidence for court proceedings, it is often hard to determine how to do this safely.

Organizations are often wary that requests for data may be coming not from clients, but from abusers. In line with their practices to reduce the threat of hostile subpoena, support workers try to "*sit the client down and talk through what's in there, and what they actually need.*" (P09). In cases where support workers can verify that a client is requesting the information, support workers further worry whether case files might hurt clients more than they would help:

> "*I'm thinking through the lens of this type of victimization. There's so much work in creating a record of what's happening for yourself, and having that kind of documentation, and pushing back against the gaslighting...it might overwhelm them.*" (P09)

Some support workers worry that case records might actually re-traumatize, or counter a survivor's healing, if reviewed without trained therapeutic support. As one said:

> "*There are things being said in there that can add to your trauma. And that's concerning, because maybe they're not in therapy anymore, so maybe they're holding onto it on their own.*" (P06)

These support workers strive to balance the benefits of providing clients direct access to their records against the potential to intensify their trauma. One widely used tactic was to supply a short and edited summary of activities that occurred in an encounter, or a "*client synopsis*" (P02) of

Phase I Finding	Use in Phase II Activity
Support workers worry whether and how to record clients' personal disclosures for service delivery, research and evaluation (4.2.2 & 4.2.3)	Part A's prompts ask survivors about their perspectives on and experiences with data collection at CETA
Case records are captured quickly, and support workers often struggle to write down enough information (4.2.1)	Part C's open records system enables users to return to a case record at any time with annotations, e.g., corrections or redactions
Clients and support workers at times disagree about whether to provide clients with their data (4.2.5)	Part B's worksheets probe each party's preferences for data sharing, to examine possible disagreement in other scenarios
Both support workers and clients feel strongly about how they are represented in these records (4.2.2 & 4.2.3)	Part C's dynamic consent is multi-stakeholder, enabling both parties to manage a case record. All activities are conducted with both parties.
Consultants feel current procedures for consent are rigid and perfunctory (4.2.4)	Part C's dynamic consent enables users to accept or reject requests for data, and update their preferences at any time

Table 2. How phase I's findings informed phase II's design provocations.

their consultation history. Composing these synopses requires further attention to the tensions in privacy and documentation reported in Section 4.2.2.

5   PHASE II: DESIGN PROVOCATIONS

Phase I's findings illustrated potential for participatory data stewardship in CETA, to give clients and consultants the control they sought over their data (4.2.2 & 4.2.3), while enabling the type of data collection needed for service delivery and funding (4.2.1). To more closely examine specific mechanisms for data stewardship in this context, we moved to phase II: a 3-part design provocation, conducted directly with IPV survivors and their support workers.

We began by priming participants to consider the data types currently collected in CETA for research (Part A, case record provocations). To more fully explore disagreements between clients and support workers in data sharing preferences (4.2.2, 4.2.3, 4.2.5), we devised a scenario-based card-sorting activity conducted with both parties (Part B, preference elicitations). Consultants' frustrations with the rigidity of the existing consent procedures (4.2.4) motivated the dynamic consent provocation; similarly, the difficulties they faced recording all relevant details in a case record (4.2.1) motivated the open records provocation (Part C, participatory mechanism provocations). Table 2 provides the full mapping of phase I findings to phase II implementations.

5.1   Methods

This phase of our study aimed to invite clients and consultants to contemplate the specific affordances of open records and dynamic consent. Methodologically, we drew on interpretive inquiry methods like design provocations and speculative data work, which have been used in CSCW to explore alternative futures for data-centric technologies [2, 12, 50, 89], and privacy preference elicitations, which have been used to understand how willing people are to share data [20]. Each of our three parts (A, B, & C, detailed below) used a provocation in the form of a mockup or worksheet to guide participants towards imagining how data is produced and circulated. Akin to Wong et al. [97], the purpose of our co-design approach was not to gather fine-grained feedback on existing

designs, but to create a space for participants to self-reflect on their values and practices around data sharing, and explore the broader politics of data stewardship in long-term use [73, 96].

We conducted our design activities over Zoom or in-person, depending on participants' preferences. Sessions lasted 60-120 minutes, depending on participants' willingness to continue discussion. Co-authors presented participants with the design exercises and provocations on a slide deck, and encouraged them to share reflections throughout. All study sessions were audio-recorded with participants' permission and transcribed using a professional service. The transcriptions were then manually anonymized by the co-authors. Interview protocols are provided in Supplementary.

Recruitment. We recruited CETA clients and consultants as participants for our design provocations. Conscious of the risks of direct research with clients of IPV services (3), we took care to minimize the potential of our research to cause retraumatization or escalation of abuse. We only reached out to clients who (a) consented to prior research participation, (b) were reachable via personal phone or email, and (c) had been seen at the clinic in the past six months. We chose the six-month window out of concerns for safety: the less recently we had seen them, the less we knew about their current situation. We then excluded clients whose participation could expose them to greater risk of harm, e.g., those who lived with their abusers. We also excluded clients whose primary language was not English, so all co-authors could conduct interviews without a translator.

Eligible clients were recruited via email. They were assured that participation would be anonymous, and would not impact their services at CETA. Clients who took part received a USD$25 giftcard as a token of appreciation. Of the 20 clients contacted, seven completed the study.[3]

Consultants were recruited via email from a random sample of active CETA volunteers. Of 11 contacted, seven completed the study, including five who participated in Phase I. Per consultation with CETA, we did not offer monetary compensation to consultants, who volunteer with the clinic.

Part A: Case record provocations. We began by priming participants to consider the data currently handled in CETA research. We gathered a list of data artifacts collected in each encounter (Table 3), and then constructed medium-fidelity mockups of a data dashboard depicting how these artifacts are stored and used in research. The data in the mockups was entirely fictional and synthetic; participants' actual clinic experiences were not displayed. Clients were prompted with these mockups to discuss their prior experiences of research at the clinic, including consent. We described when and how each piece of information is collected, and asked how well participants felt their clinical encounters were captured by the data.

Part B: Preference elicitation activities. Next, we explored participants' high-level preferences for sharing their data. To scaffold these discussions, we used a modified card-sorting exercise, akin to Markussen and Knutz's *mini-scenarios* [62]. Though card sorting originated as a technique in informatics to understand ontology, prior work has used it in studies of privacy preferences, to help distill complex concepts into concrete artifacts that encourage participants' imaginations [20].

Inspired by Nissenbaum's parameters of information flow [66], we used card sorting to separate different stated *purposes* for data sharing from the different social roles who might be *recipients* of the data (e.g., separating the purpose of *raising public awareness of tech abuse* from a recipient such as *journalists and the press*). This was to understand how participants' desired uses of their data might differ from what they believed people might do with it. Drawing from data requests CETA has received, we derived six potential recipients and five potential purposes (Table 3).

We used these artifacts, recipients and purposes in a preference elicitation activity centered on a worksheet (Figure 2). Each worksheet contained a specific purpose or recipient, a list of

---

[3]This response rate is higher than typical in IPV research: since clients of IPV services often change their contact methods and desire to "move on", they have low rates of re-engagement in research [93].

Data artifacts	Recipients	Purposes
Your/Client's age, gender, race / ethnicity	Journalists / Press	To make laws
Transcripts of your appointments	Technology Companies	To raise public awareness of tech abuse
Notes from your appointments	Policymakers	To improve services like CETA
Your/Client's experiences with tech abuse	Law Enforcement	To punish abusers
Your/Client's referral forms	Academic Researchers	To build technology products
<Blank - fill in>	Advocacy Communities	

Table 3. Data artifacts and recipients used in the design activities. Each artifact (except the last one) is currently collected in CETA research. Recipients include specific *social roles* to which data might be released. We separate these from high-level *purposes* for its sharing. Using the worksheet in Figure 2, participants were asked to relate their preferences for sharing data to combinations of artifacts and recipients.

Fig. 2. Worksheet used in Part B. Each worksheet includes a target recipient/purpose (see Table 3) and a list of data artifacts (middle). For each artifact, participants move its card onto the scale according to their level of comfort sharing it with the recipient/purpose (bottom), or to the "I do not want to share" box (top right).

cards representing CETA data artifacts, and a scale ranging from "least comfortable sharing" to "most comfortable sharing", with five markers in between. Participants were asked to drag and drop the cards onto the scale according to their data sharing preferences. The worksheet also contained a "do-not-share" area for any artifact participants were not comfortable sharing under any circumstances—providing space for 'do-not-design' implications reflecting when data should *not* be shared at all [78]. Participants were encouraged to use the scale flexibly (e.g., to stack items at the same marker or place items between markers), to add and modify data artifacts as needed, and to talk-aloud, describing their doubts, questions, and rationales.

Part C: Participatory data stewardship provocations. Having primed participants to consider the data collected at CETA and how they would like it shared, we then explored the possibilities of our chosen mechanisms for participatory data stewardship: *open records* and *dynamic consent*.

We extended the case records mockups from part A into a mockup of an open records system, in which a client or consultant may view and comment on CETA's records of their clinical encounters (Figure 3). Participants first considered a *tagging* feature, highlighting the process of categorization
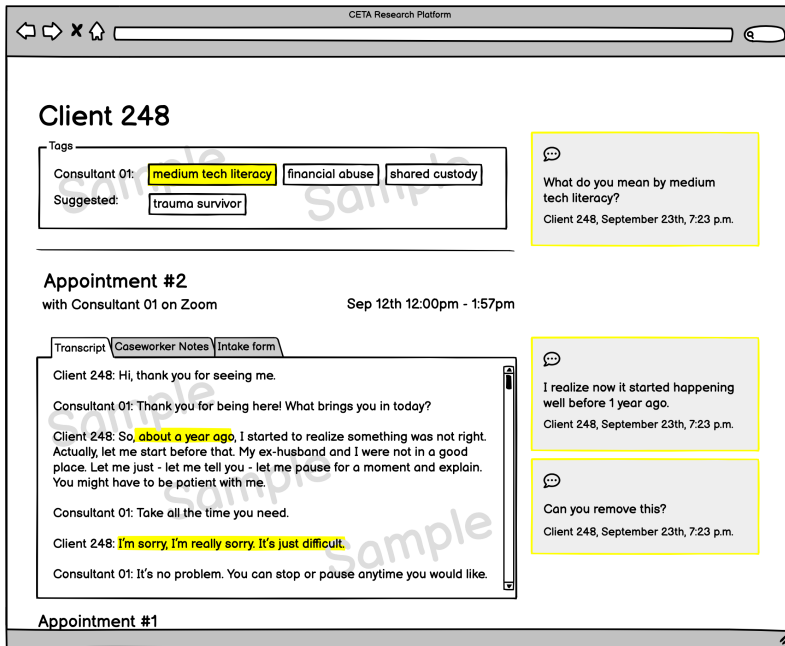
Fig. 3. Medium-fidelity software mockup of an open record (phase II, Part C). The fictional Client 248 can view tags applied to their case record by consultants (top), and annotate to suggest edits or redactions (right).

that underlies research. We included tags that could be perceived as subjective or personal (e.g., *medium tech literacy* or *shared custody*) to gauge participants' comfort with different types of research. Tags were sourced by consultants and suggested by artificial intelligence (AI). Next, participants considered a *commenting* feature, which enables annotations on a case record in the style of tools like Google Docs (e.g., the fictional Client 248 commenting "*Can you remove this?*").

The next mockups depicted a system for *dynamic consent* (see Supplementary). These affordances included receiving requests for data from external researchers, approving or denying release of their data, and subscribing to updates on approved projects. Participants could choose to partially consent to requests, by releasing some data, but not all (e.g., releasing demographic information, but not appointment transcripts). Participants were prompted to reflect on the utility of dynamic consent for their CETA data, and if/how they could imagine engaging with incoming requests.

We concluded sessions by asking participants to imagine a scenario in which they were not available to respond to incoming data requests. Questions probed who they would trust to make these decisions on their behalf, and what they would like that person or organization to consider.

Analytical approach. We conducted design sessions with seven clients and seven consultants. Session transcripts and notes were analyzed using a reflexive thematic analysis adapted from Braun & Clarke [18]. Across stakeholder groups, we independently coded transcript data and developed an initial set of 30 codes (e.g., "*giving clients control, open records shape memory, accountability over punishment*"). These codes were further honed into themes through multiple rounds of collaborative

refinement with the research team. Finally, we tabulated the data from the preference elicitation activities to develop the qualitative heatmap depicted in Figure 4.

Ethics. We took the same approach as Section 4.1 with additional steps to protect participant anonymity. Clients may still be coping with abuse and revealing they are seeking support may inadvertently risk escalation of abuse. To ensure participant safety, we refrain from reporting demographic data and removed potentially identifying details from our findings and provocations.

We also took care to minimize risks that our research could impose on participants' health. Design activities were conducted by authors trained in trauma-informed research. We ensured participants knew they could pause or stop the session, decline to answer questions, or withdraw from the study completely. For clients, we made clear that research activities were separate from CETA services, and ensured their contact points for research and service did not overlap. We also worked with CETA to anticipate requests for additional service from research participants. Lastly, for participants who chose an in-person session, we used a location with on-site security guards.

5.2 Findings

Our findings (Table 1) illuminate the design space around participatory mechanisms for data stewardship at CETA. Here, we refer to client participants with the letter S and their participant number; and likewise, consultant participants with the letter C. Where we refer to one subgroup, we use either *client(s)* or *consultant(s)*; where we refer to both subgroups as one, we use *participant(s)*.

5.2.1 Clients also consider informed consent to be rigid and perfunctory. Part A of our sessions asked clients to reflect on their experiences of giving informed consent to CETA research. At the time of our study, CETA's consent process had been a verbal consent delivered over a phone call at the start of an appointment. Clients are informed of the study and the associated risks and benefits of participating, after which they either affirm or refuse participation verbally. [4]

All 7 participants said their experiences of consent barely left an impression, but to their memory, the process had been brief and clear. They were familiar with consent forms, from the media or from research ethics trainings in their education or employment. As one explained, they said yes to CETA to help accumulate "accurate" data, for broader research goals:

> "*I assumed [I agreed] just to help whatever you got a grant to research. Probably the more people say yes, it probably just helps the pool of data be more accurate.*" (S06)

This finding affirms our finding in 4.2.4: existing methods of obtaining informed consent are considered rigid and perfunctory. We unpack this further in 6.2.1 and 6.2.2.

5.2.2 Open records could provide self-reflection and control over survivor narratives. All participants had generally positive reactions to the idea of giving clients access to their own case records (Part B). One key potential benefit was enabling clients' **self-reflection**. Clients envisioned using these tools to track how their digital privacy skills developed over time, assess any need for additional support, and bolster their healing. As one said: "*It would be wonderful, because you can teach someone how to protect themselves, and they can learn from it.*" (S01)

Clients also saw utility in **annotating** their records, via the commenting affordance in Figure 3. During appointments, they sometimes had trouble describing their complete experiences, due to trauma's negative impact on memory:

---

[4]The verbal consent was delivered instead of the more traditional paper form as a safety measure, to prevent inadvertent harms from abusers discovering consent forms that said a client had participated in a study about abuse.

"*Sometimes in those situations, we may forget something or not be as accurate with it. I know when I did my appointment, being that I was still kind of in the midst of [the abuse], my mind was just all over the place. I had so much stuff going on.*" (S04)

With open records, however, clients imagined logging into the provocation once or twice, weeks to months after a clinical encounter, to make sure an accurate representation was recorded. These records could then help establish **documentation** of abuse, for use in legal proceedings.

Still, clients and consultants both said there might be moments where a client would want access to their record to **redact** information, instead of improving its accuracy. Redaction would be particularly vital for clients whose abusers could resurface after a long period of time. One such client wanted to remove any mention of her location from her records:

"*As an extra measure of precaution, I would want to make sure if I had shared something like my address, I could take that off. These people can be really vindictive. I would want to know that I did everything I could to protect myself and my children.*" (S05)

Intriguingly, clients were divided on whether or not removing information from a clinical record would be ultimately helpful. While providing clients opportunities to regain control of their lives, these redacted records could be unrecoverable for later use in court proceedings or in self-reflection.

Whether for annotation or redaction, clients agreed they could see themselves making use of the provocation to **exert control over their narratives** and reclaim agency. Decisions about how to manage their records could, in turn, be spaces for additional self-reflection. One client said their desire to take things on or off the record could be material for therapy:

"*You with your therapist or doctor can come back with your questions, and you guys can work on something. Like, "why do you want to remove this? Is it too emotional? It's real. It is what you experienced. Let's talk about that."*" (S03)

This finding contrasts with 4.2.5, in which support workers expressed access to case records may re-traumatize clients. We unpack this discordance further in 6.1.2.

Out of all the affordances in Part A, clients and consultants were most interested in how the **tagging** feature might enable service delivery, broader research, and sense-making of clients' experiences. Consultants suggested tags for notable cases, like ones involving firearms or children, could help understand trends in client referrals. One client said the tags could support healing, by enabling survivors to identify as a victim, a survivor, or neither:

"*Those are actually awesome tags. Like 'trauma survivor'. Not 'abused', because they could be a survivor instead of being seen as a victim . . . Maybe you could also have the option for 'victim' as well, though, because some people haven't yet felt like a survivor. They just feel like a victim.*" (S02)

But clients were also concerned that tags could become labels, which if handled improperly could reduce trust between clients and their support workers. Highly sensitive labels, like "*shared custody*", are often useful for support workers, who can better safety plan if they know children are going between the client's and abuser's households. However, concretizing such painful situations on the record could do further harm to clients. As one client said:

"*[The tags] might just turn me off, and then I don't want to talk further. Maybe it puts salt on a wound that is even bigger, and seeing it stated like that hurts me deeply.*" (S01)

To our participants, appropriate use of the tags meant enabling clients to use them in self-reflection and dialogue with support workers, while avoiding further harms from labeling clients with traumatic events—or narratives dissonant with their own. We return to this point in 6.1.2.

5.2.3 Granular selective sharing in dynamic consent accommodates diverse preferences. Clients and consultants saw that the dynamic consent tool (Part C) could provide greater information control,
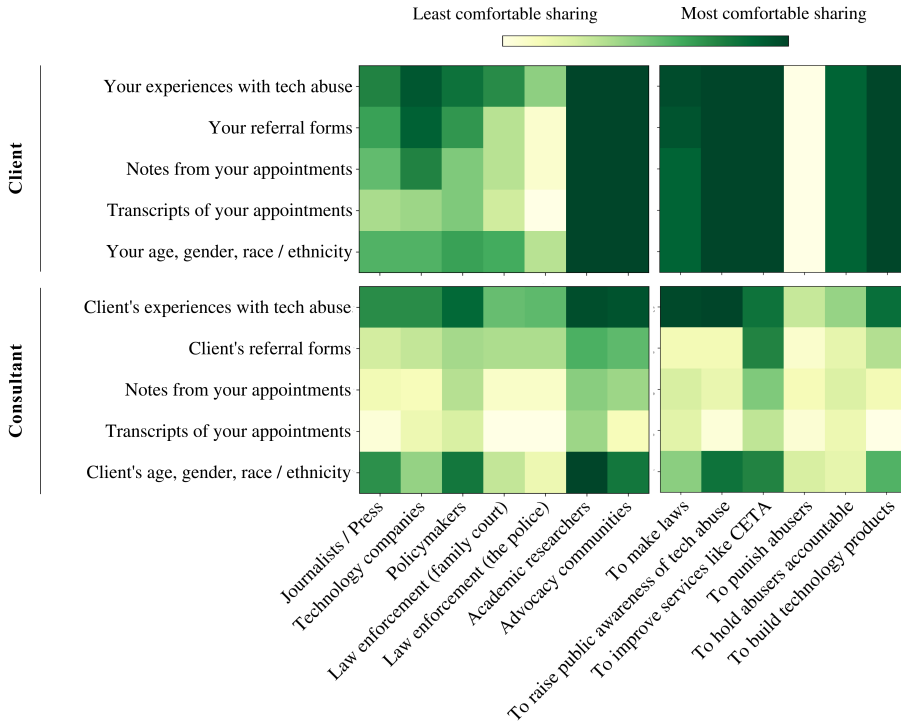
Fig. 4. Participants were asked to rate how comfortable they felt on a scale from 0-5 with sharing different data artifacts from CETA encounters (y-axis) to various social roles or purposes (x-axis). Exact numbers are omitted here due to the qualitative nature of exercise, but are available in Supplementary.

in a digital society where they felt like they had less. Recalling perfunctory experiences of consent (5.2.1), one client said dynamic systems could offer much-needed transparency and trust:

"*I feel like a lot of times our data is collected, and the terms of use are in this tiny font where you accept, and you don't read it because you're like, I'm not going to sit here and read this. But your data is collected, and nobody explains what it's used for, or why. So offering transparency, I feel would just build trust in your organization.*" (S06)

Beyond transparency, clients and consultants appreciated the opportunity to give control of data back to the subject—especially given the nature of IPV. The ability to retract information shared to a third party was seen as a key way to react to changing safety situations. As one client described:

"*I want to share my experience, but God forbid something happens, and now I want to change my mind and protect myself. It shouldn't be that once it's out there ...I can't take it down.*" (S05)

Both clients and consultants agreed that **selective sharing**, or the ability to share some data with some recipients and not others, would be critical to implementing dynamic consent. Importantly, selective sharing needed to be individual: participants varied widely in their data sharing preferences, and consultants were far less permissive than clients (Figure 4).

Many clients expressed enthusiasm for the possibility of their data being used to improve society by spreading awareness of tech abuse. As one client said: "*Put it all in 'most comfortable sharing.'*

Maybe someone will listen and make change." (S01). Another client shared that while she was initially afraid to share her experiences, she was inspired to contribute to a broader movement:

"*Not everybody is comfortable with sharing very private and personal things. But sometimes, when you're brave, you share things to help someone who isn't as brave as you are. And sometimes it has, like, a #MeToo kind of effect, where other people come out of the woodwork and share their experiences. When you keep things to yourself, it keeps happening, and you feel isolated. But when you let it be known, we can heal and move forward and make a type of movement.*" (S02)

In spite of this enthusiasm for collective gains, when asked to consider specific mini-scenarios for their own data, clients had hesitations. Even within our cohort of 7 people, clients *differed widely* in their data sharing preferences. Some felt tech companies could not be trusted with their data, as it would be leaked, or used in behavioral advertising. Others felt letting tech companies know about tech abuse would help them improve their products' robustness against it. Some trusted journalists and lawmakers implicitly, while others only trusted those who aligned with their politics. Most trusted academics and advocates—though we must be aware of sample bias limitations (6.3).

Our clients were least comfortable sharing data for the purpose of punishing abusers, or to roles in law enforcement, like family court or the police (Figure 4). While clients generally wanted to see consequences for abusers, some felt the criminal legal system was a form of unnecessary cruelty that "*did not create a better space for anybody*" (S03). Importantly, several clients were hesitant to share data with the police, after previous experiences with officers dismissing their abuse. For some, law enforcement's dismissiveness was a *reason* to share data—to provide incontrovertible proof of what happened. For others, this was a reason to avoid the police altogether.

As an alternative, clients suggested lending data to boost efforts at restorative justice. Several asked to modify the scenario to say "*to hold abusers accountable*", instead of "*to punish abusers*":

"*If we are holding people accountable, and we're able to do that in a safe space with this information, I think this would be very helpful. How do you hold someone accountable if you don't know what's happening—the weight or gravity of it, or the impact it's having on victims?*" (S03)

CETA consultants, for their part, were generally less willing than clients to share case records (Figure 4). Having also trained to advise on digital privacy, they defaulted to more protective security postures. For example, many wanted to modify the artifacts to contain less raw information (Figure 4), and instead share curated written summaries, paraphrased quotes, or aggregated statistics.

Consultants particularly worried about secondary use of demographic data. Many raised that the intersection of demographic attributes could be re-identifying for individual clients. While aggregated demographic data might inform efforts towards "*diversity, equity, inclusion, all that good stuff*" (C07), in the wrong hands it could also inspire targeted discrimination, "*to fuel hate crimes*" (C02). We discuss the variance in sharing preferences further in 6.1.2 and 6.2.2.

5.2.4 Alongside granular and self-directed information control, a role remains for trusted stewards. Despite the benefits they saw in open case records and dynamic consent, our participants were clear: there remained a role for a trusted steward to ease the burden of managing client data. After completing the preference elicitation worksheets (Part B), clients and consultants alike said these decisions were challenging, requiring time and effort to consider satisfactorily. Affirming what professionals and consultants surmised in 4.2, several clients said the burden of managing their case records had the potential to re-traumatize. As one described:

"*The more people can learn from my experience, the better. But I also don't want to be overwhelmed. I don't want to be interviewing with a bunch of people, telling this story over and over again.*" (S04)

Trusting a data steward was seen as a way for clients to contribute to the collective goals of IPV research while mitigating their own burdens. A key role for this person or entity was to monitor and

vet the researchers requesting the data. Participants were wary of the burden that handling requests for data would incur in their lives, despite the potential benefits. Some wanted a notification every time a request came in; others wanted weekly or monthly digests. Several consultants did not want to manage incoming requests at all, and deferred this responsibility entirely to a steward. All said their preferences likely varied from others'—one client mused that the only solution might be to have individual data subjects set their own preferences for how often they receive requests.

In addition to judiciously meting out how often data subjects received requests, participants asked that a data steward take an active role in vetting incoming requests, so they would not have to research whether requests were problematic or illegitimate. At issue was what one client described as "*integrity*" (S04)—could the requester be trusted to use the data only for its intended purpose? (We unpack the possibility of enforcing intended use in 6.2.2.)

In addition, several clients asked that stewards consider the "*efficiency*" (S04) of the data requesters. Receiving research updates would be very interesting, these clients said, to see whether the people trusted with their data used it properly. Proper use meant not only whether the data was used only for intended purposes, but also whether it was used in a timely fashion. As one client said:

> "*You can collect data for years and years and years, and you don't see the purpose. So I would like to see something so we can know that we're not just basically wasting our time.*" (S05)

As for who the steward should be, some participants named family members, like siblings or parents—but most said CETA itself, because they were already familiar with the organization. Several clients said they would trust the specific consultant responsible for their case. Consultants, for their part, said they trusted CETA as a nonprofit organization to put clients' best interests first.

6 DISCUSSION

Our findings (Table 1) document our efforts to co-design mechanisms for participatory data stewardship in a frontline IPV support service, to improve survivors' agency in the knowledge produced through these encounters. Overall, we are encouraged by the potential benefits identified in our work: creating spaces for self-reflection and self-determination for clients, and encouraging deeper research in IPV that helps organizations achieve funding and impact. For these reasons, we encourage pursuit of participatory data stewardship in clinical contexts, in IPV and beyond.

Realizing this potential will, however, require further work adapting participatory data stewardship for the challenges of digital-safety research. Here, we extend the theoretical framework for participatory data stewardship synthesized in Figure 1 by analyzing the tensions our work uncovered in their use (6.1). We then describe practical next steps for the growing research community interested in participatory approaches to digital safety (6.2), and close with limitations (6.3).

6.1 Extending participatory data stewardship for digital-safety research

Our work complicates a core contention of current frameworks for participatory data stewardship: that *collective* rather than *individual* mechanisms for data management are more capable of enabling meaningful participation. Because power imbalances make individuals "*rarely in a position to negotiate or engage in trade-offs as they relate to their data rights*" [1], these frameworks argue that the goal of data stewardship is to enable a collective to make decisions about its constituents' data (Figure 1, top-half), and reduce researchers' supervision over the process (Figure 1, top-left). We find, however, that to apply this to digital-safety research, we need further work reconciling: (1) tensions between individual vs. collective benefit in data stewardship; and (2) the need for nuanced collaboration between researchers and data subjects, despite inherent power imbalances.

6.1.1 Individualized mechanisms for participation still hold value in digital safety. The Lovelace framework critiques individualized mechanisms for data handling, such as open records and

dynamic consent, for failing broader collective and scientific goals. Since individuals can remove their data at will, the stewarded data can become inaccurate: e.g., government-hosted biobanks are known to under-represent minority populations, who are often less comfortable donating their data [1]. These issues are part of why the literature encourages community governance, e.g., a committee of IPV survivors owning an archive of IPV stories. With community supervision, individual data subjects and researchers can both be held accountable to the broader scientific aims of the project.

Such an approach, however, may overlook pragmatics in IPV and digital safety. First, we find open records and dynamic consent have unique utility for data subjects facing digital-safety threats: opportunities for self-reflection and skill-building that lead to feelings of empowerment (5.2.2), a key goal in digital safety (cf. [88, 90]). Second, the idea of community governance may be at odds with the practical realities of IPV. Our data show clients of an IPV support service may not identify in uniform ways, making it hard to identify which collectives should represent them: as one said, "*some people haven't yet felt like a survivor. They just feel like a victim.*" (S02, 5.2.2). What's more, even among the seven in our study, clients disagreed about appropriate uses of their data (Figure 4). While some saw merit in "*making a type of movement*" (S02, 5.2.3), abstracting an individual survivor's decisions to a collective might disempower the individual, or even risk their safety.

To realize the benefits of participatory mechanisms for data stewardship in digital-safety research, we argue it is important to consider that a survivor's individual empowerment may at times be as important as rallying with a collective. In many cases, supporting the individual over the potential of collective benefit may be the preferred option, or the only feasible one. Further work is needed to fully interrogate what constitutes meaningful participation for individuals, and how to balance individual versus collective gains in the stewardship of digital-safety data.

6.1.2 Refocusing from researcher vs. community control to researcher-community collaboration. Frameworks for participatory data stewardship describe a dialectic: between researchers as the power-holders, and the data subjects and community as the powerless or vulnerable. Both Lovelace [1] and Jo and Gebru [53] frame the problem of participation as oscillation between whether the researchers or the data subjects are in charge. Our findings instead suggest that in digital safety, it is important to understand data stewardship systems as *collaborations* between researchers and data subjects. This is *not* to say that researchers and data subjects are equal parties; our findings support work on the uneven power dynamics inherent to both research and service delivery in IPV (cf. [10, 90]). Rather, researchers and subjects are both relevant stakeholders—in the work of stewarding data and in narrative control over what the data represent.

First, we identify an important role for an *active* steward in the effort required to safely manage digital-safety data. Figure 1 shows how increased community supervision can also levy burdens on the community. To reduce these burdens, prior work in citizen science [69] has argued that researchers should be *facilitators*, supplying technical expertise in areas like data security, but otherwise taking a backseat to community control. In contrast, survivors in our study expressed a desire for stewards to help *make decisions* about their data: to take an active role in managing this information ad infinitum (5.2.4). To make these decisions, stewards cannot rely only on their own judgment: support workers, who act as stewards in this context, tend to be more cautious than survivors (Figure 4). In situations where the two parties disagree, a client-centered approach may suggest favoring clients' wishes and releasing the data. But doing so requires stewards to then help clients make sense of their information (4.2.5), and in the worst case, may implicate stewards' own safety (4.2.3). These tensions are not resolvable by giving a collective of survivors total supervision of the archive; instead, the burdens of supervision must be handled via collaboration and compromise. Future work should explore how an active steward, collaborating *with* a community, might help disentangle unwanted burdens from the desired benefits of community supervision.

More broadly, our findings also suggest stewards have an active stake in ownership over *the data itself*. From the notes written to the tone of the conversation, a clinical encounter is a co-production between a client and a support worker—even when it seems only a client is disclosing personal details (4.2.1). These exchanges should not be understood as static capture of clients' experiences, but rather sites of collaborative sensemaking. Support workers are, in a sense, also data subjects—their beliefs, experiences, and intellects are also reflected in the data. When they also serve as researchers, as in our case, and thus as stewards, they have an even greater stake in the data produced. We see ample future work building on this observation to reconstruct data stewardship as a multi-stakeholder system, accounting for multiplicity in the dynamics between researchers, survivors, and their communities.

This perspective begs a broader question: What happens when survivors' self-narratives clash with the narratives that support workers or researchers might see in their stories? As survivors told us in reaction to the tagging feature of our provocation (5.2.2), seeing one's life characterized via painful events or dissonant narratives can be harmful. Just as studies of medical misinformation must contend with dissonant narratives [59], so too must participatory stewardship of digital-safety data balance experience, belief, expertise and power in making sense of human experience. Stewards should not prohibit research on important digital-safety problems if it might cause friction with clients' self-narratives; but they must still preserve clients' autonomy of refusal. Turning such dissonance into productive collaboration will require moral and political judgment, but it will be crucial to reaping the benefits of these systems across digital-safety contexts.

6.2 Practical recommendations for researchers

In addition to the theoretical advancements suggested above, our work suggests practical future work exploring the design space around participatory data stewardship for digital-safety research.

6.2.1 Improving informed consent. Informed consent is undeniably integral to ethical research [56], especially with people like IPV survivors [93]. It is also critical to clinical practice, and to stewarding data in FAIR ways: for data to be shared, it must be collected ethically in the first place.

However, our findings underscore the need for new implementations of consent. The traditional model—of upfront contractual agreement to a written form or verbal script—has become so familiar that clients give a perfunctory "*yes*" (4.2.4 and 5.2.1). To be clear, perfunctory agreement does not mean clients are not informed. Several said they breezed quickly through consent *because* it was already very familiar, and all said they were comfortable giving data to CETA. But our findings show informed consent has become so familiar a procedure that it risks becoming just a checkbox.

In light of our findings, CETA has begun adapting its consent processes, making changes away from starting appointments with a "*wall of text*" and "*legalese*" (4.2.4) and towards a more conversational style. This approach aims to uphold survivors' existing skills at negotiating safety risks [24]. Paired with the tools for dynamic consent explored in Section 5, this adjustment stands to improve how research and service is conducted in clinical settings. We look forward to future work joining a growing community in CSCW and HCI exploring consentful interfaces (cf. [52, 98]).

6.2.2 Roles for computation in participatory data stewardship. Downstream of informed consent, we also need mechanisms for how a data steward can *feasibly* manage a growing and ever-changing digital-safety dataset, as well as requests for re-use from secondary researchers.

A steward might first need tools ensuring shared data are appropriately sanitized. Here, what counts as sensitive is highly contextual: support workers use subjective judgment in minimizing the data they record (4.2.2 and 4.2.3), and clients and consultants differ in their privacy preferences (5.2.1). Thus, existing sanitization techniques like algorithms that search for names or emails are

insufficient, especially since the data are unstructured and multimodal. Elicitation tools like the worksheets in our study (5.1) may help stewards understand what data subjects consider sensitive.

A steward might then need tools for matching incoming requests for data to case records. This is needed to rate-limit requests that come to clients and consultants, thereby easing the burden of dynamic consent (5.2.4): if a research question has little relevance to a case, and the client or consultant has expressed disinterest in similar research, there is less reason to consider it. Matching requests to relevant records will require abstractions beyond keyword searches, as researchers and stewards may not know *a priori* how a construct manifests in the data—for example, how can a researcher query for expressions of trauma before knowing how it is expressed? Stewards will need to match abstractions from requests to the abstractions of preferences described previously.

Finally, stewards might need tools enforcing accountability for the research conducted with the shared data. These tools might vet requesters for integrity and efficiency (5.2.4), and help ensure they stick to the approved purposes of the data. Importantly, stewards must consider how to enable data subjects to redact data, while ensuring researchers still receive sufficient information.

At each of these steps, computational tools like semantic search might help stewards accomplish these goals, while reducing their own burdens—however, these tools will need robust human oversight. We see future work building complementary human-AI systems towards these needs, akin to efforts to use ML to help activists document feminicides [30].

6.2.3 Making stakeholder-guided design sustainable.

As discussed in the Lovelace framework, mechanisms for participatory data stewardship are best developed via co-design: journeys of incremental design and deployment, requiring frequent and meaningful involvement from stakeholders.

Such work has proved costly. Inviting stakeholders to reimagine how their data is handled required us to ensure our study did not itself confer additional burdens on clients, consultants, or professionals. We considered that our clients had experienced trauma (cf. [23]); that care workers are exposed to secondary trauma every day; and that power differentials within and between organizations shape how participants interact with us. We also considered that participants face societal-scale problems that may not be wholly solvable, let alone via technology, and inviting them to co-design might risk Harrington et al.'s [46] "blue-sky" ideation: encouraging imagination of infeasible solutions that "*ultimately frustrate underserved individuals*". The burden of participation here constituted not only time and effort, but also what Dourish et al. [33] describe as the *emotional labor* of hoping and believing one can change their structural condition.

These tensions manifested in a study whose design and execution required us as researchers to also act as stakeholders in the co-design process, making judgment calls that structured each encounter. Each session took considerable effort to maintain what Akama and Light call punctuation and poise [3]: orientations to research where researchers stay ready to handle contingencies outside of a dogmatic march through a protocol. In the phase II activities (5), for example, some sessions lasted more than three times the planned 60 minutes, as researchers gave the participant ample room to recount their abuse before returning to the interview guide. That design research encounters can resemble therapy has been discussed previously in HCI and CSCW, where Hirsch has observed that the obligations of a design researcher are quite different from those of a therapist [49]. Yet, in our context, we found it especially difficult to separate our obligations as researchers from our inclinations to provide support, and took extra steps to coordinate follow-up care with CETA for participants who sought it. There was, too, the issue of safety: we ensured our interview locations were reasonably secure, for researchers and participants alike who face threats from abusers.

We are confident the extra precautions we took were necessary, and improved the depth of our work. We raise them here to do justice to our participants, and to researchers' well-being. How to make such deep participatory engagements *sustainable* is a critical question facing future

work. We call for greater attention in CSCW to methodological standards that can help researchers more efficiently navigate co-design with vulnerable participants. As starting points, CSCW could incentivize documenting and publishing on the emotion work [7] that such research requires of stewards and researchers; and of participants and their broader communities.

6.3 Limitations

This was a single-site study examining participatory data stewardship in CETA's ecosystem in New York City. Further work is needed to examine these possibilities in other clinical settings that serve other populations. Multi- and cross-site research may more fully explore how race, gender, and other social factors influence participants' perspectives.

Importantly, our provocation exercises (5) should be understood as *early explorations* of design considerations for participatory data stewardship in digital safety, not generalizable statements on population-wide preferences, or acceptance testing of new modalities. This was a research study of people's willingness to participate in research—so our results do not characterize those who would not have consented to work with us at all. Lastly, our work is reflexive, and thus a product of our lenses. Future work might take an alternative approach, like broad surveys of data subjects' preferences, or ethnographic engagements by researchers with no role within the site. These approaches, combined with ours, will provide richer explorations of these possibilities.

ACKNOWLEDGMENTS

REFERENCES

[1] Ada Lovelace Institute. 2021. *Participatory Data Stewardship: A framework for involving people in the use of data.* Technical Report. https://www.adalovelaceinstitute.org/report/participatory-data-stewardship/

[2] Daniel A. Adler, Emily Tseng, Khatiya C. Moon, John Q. Young, John M. Kane, Emanuel Moss, David C. Mohr, and Tanzeem Choudhury. 2022. Burnout and the Quantified Workplace: Tensions around Personal Sensing Interventions for Stress in Resident Physicians. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (Nov. 2022), 1–48. https://doi.org/10.1145/3555531

[3] Yoko Akama and Ann Light. 2018. Practices of readiness: punctuation, poise and the contingencies of participatory design. In *Proceedings of the 15th Participatory Design Conference: Full Papers-Volume 1.* 1–12.

[4] Tim Althoff, Kevin Clark, and Jure Leskovec. 2016. Large-scale analysis of counseling conversations: An application of natural language processing to mental health. *Transactions of the Association for Computational Linguistics* 4 (2016), 463–476.

[5] Sherry R Arnstein. 1969. A ladder of citizen participation. *Journal of the American Institute of planners* 35, 4 (1969), 216–224.

[6] Karen S Baker and Helena Karasti. 2018. Data care and its politics: Designing for local collective data management as a neglected thing. In *Proceedings of the 15th Participatory Design Conference: Full Papers-Volume 1.* 1–12.

[7] Madeline Balaam, Rob Comber, Rachel E Clarke, Charles Windlin, Anna Ståhl, Kristina Höök, and Geraldine Fitzpatrick. 2019. Emotion work in experience-centered design. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* 1–12.

[8] Solon Barocas and Helen Nissenbaum. 2014. Big data's end run around anonymity and consent. *Privacy, big data, and the public good: Frameworks for engagement* 1 (2014), 44–75.

[9] Rosanna Bellini. 2023. Paying the Price: When Intimate Partners Use Technology for Financial Harm. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '23).* ACM, New York, NY, USA. event-place: Hamburg, Germany.

[10] Abeba Birhane, William Isaac, Vinodkumar Prabhakaran, Mark Diaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed. 2022. Power to the people? opportunities and challenges for participatory AI. *Equity and Access in*

Algorithms, Mechanisms, and Optimization (2022), 1–8.

[11] Guthrie S Birkhead, Michael Klompas, and Nirav R Shah. 2015. Uses of electronic health records for public health surveillance to advance public health. *Annual review of public health* 36 (2015), 345–359.

[12] Kirsten Boehner, Janet Vertesi, Phoebe Sengers, and Paul Dourish. 2007. How HCI interprets the probes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. Association for Computing Machinery, New York, NY, USA, 1077–1086. https://doi.org/10.1145/1240624.1240789

[13] Sarah R. Boonin. 2016. Ten Years Too Long - Reforming Social Security's Marriage Duration Requirement in Cases of Domestic Violence. *Harvard Journal of Law and Gender* 39, 2 (2016), 369–428. https://heinonline.org/HOL/P?h=hein.journals/hwlj39&i=383

[14] Christine L Borgman. 2012. The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology* 63, 6 (2012), 1059–1078.

[15] Geoffrey C Bowker and Susan Leigh Star. 2000. *Sorting things out: Classification and its consequences.* MIT press.

[16] danah boyd. 2022. Crisis Text Line, from my perspective. https://www.zephoria.org/thoughts/archives/2022/01/31/crisis-text-line-from-my-perspective.html

[17] danah boyd and Kate Crawford. 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society* 15, 5 (2012), 662–679.

[18] Virginia Braun and Victoria Clarke. 2021. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology* 18, 3 (July 2021), 328–352. https://doi.org/10.1080/14780887.2020.1769238

[19] Charles L Briggs. 1986. *Learning how to ask: A sociolinguistic appraisal of the role of the interview in social science research.* Number 1. Cambridge University Press.

[20] Kelly Caine and Rima Hanania. 2013. Patients want granular privacy control over health information in electronic medical records. *Journal of the American Medical Informatics Association* 20, 1 (Jan. 2013), 7–15. https://doi.org/10.1136/amiajnl-2012-001023

[21] Jacquelyn C. Campbell and Jennifer Manganello. 2006. Changing Public Attitudes as a Prevention Strategy to Reduce Intimate Partner Violence. *Journal of Aggression, Maltreatment & Trauma* 13, 3-4 (Nov. 2006), 13–39. https://doi.org/10.1300/J146v13n03_02 Publisher: Routledge _eprint: https://doi.org/10.1300/J146v13n03_02.

[22] Centre for Disease Control and Prevention. 2020. Fast Facts: Preventing Intimate Partner Violence.

[23] 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 *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–20. https://doi.org/10.1145/3491102.3517475

[24] Rachel Clarke, Peter Wright, Madeline Balaam, and John McCarthy. 2013. Digital Portraits: Photo-sharing After Domestic Violence. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 2517–2526. https://doi.org/10.1145/2470654.2481348 event-place: Paris, France.

[25] Allison Crowe and Christine E. Murray. 2015. Stigma From Professional Helpers Toward Survivors of Intimate Partner Violence. *Partner Abuse* 6, 2 (Jan. 2015), 157–179. https://doi.org/10.1891/1946-6560.6.2.157 Publisher: Springer Section: Research.

[26] Dana Cuomo and Natalie Dolci. 2021. New tools, old abuse: Technology-Enabled Coercive Control (TECC). *Geoforum* 126 (Nov. 2021), 224–232. https://doi.org/10.1016/j.geoforum.2021.08.002

[27] Tom Delbanco, Jan Walker, Jonathan D Darer, Joann G Elmore, Henry J Feldman, Suzanne G Leveille, James D Ralston, Stephen E Ross, Elisabeth Vodicka, and Valerie D Weber. 2010. Open notes: doctors and patients signing on. , 121–125 pages.

[28] Sandi Dheensa. 2020. *Recording and sharing information about domestic violence/abuse in the health service: Research report and good practice recommendations for healthcare.* Technical Report. Pathfinder consortium, Bristol University. 1–113 pages. https://research-information.bris.ac.uk/ws/portalfiles/portal/248068597/FINAL_REPORT_Recording_and_sharing_DVA_information_in_healthcare.pdf

[29] Catherine D'Ignazio. 2022. *Chapter 8 – A Toolkit for Counterdata Science.* https://mitpressonpubpub.mitpress.mit.edu/pub/cf-chap8.

[30] Catherine D'Ignazio, Isadora Cruxên, Helena Suárez Val, Angeles Martinez Cuba, Mariel García-Montes, Silvana Fumega, Harini Suresh, and Wonyoung So. 2022. Feminicide and counterdata production: Activist efforts to monitor and challenge gender-related violence. *Patterns* 3, 7 (July 2022), 100530. https://doi.org/10.1016/j.patter.2022.100530

[31] Karthik Dinakar, Allison JB Chaney, Henry Lieberman, and David M Blei. 2014. Real-time topic models for crisis counseling. In *Twentieth ACM Conference on Knowledge Discovery and Data Mining, Data Science for the Social Good Workshop*. Citeseer.

[32] Lynn Dombrowski, Ellie Harmon, and Sarah Fox. 2016. Social justice-oriented interaction design: Outlining key design strategies and commitments. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. 656–671.

[33] Paul Dourish, Christopher Lawrence, Tuck Wah Leong, and Greg Wadley. 2020. On being iterated: The affective demands of design participation. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–11.

[34] 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 (July 2018), 609–625. https://doi.org/10.1080/14680777.2018.1447341 Publisher: Routledge _eprint: https://doi.org/10.1080/14680777.2018.1447341.

[35] Stephanie OM Dyke, Kathleen Connor, Victoria Nembaware, Nchangwi S Munung, Kathy Reinold, Giselle Kerry, Mamana Mbiyavanga, Lyndon Zass, Mauricio Moldes, Samir Das, et al. 2022. Consent Codes: Maintaining Consent in an Ever-expanding Open Science Ecosystem. *Neuroinformatics* (2022), 1–12.

[36] Paul N Edwards, Steven J Jackson, Melissa K Chalmers, Geoffrey C Bowker, Christine L Borgman, David Ribes, Matt Burton, and Scout Calvert. 2013. Knowledge infrastructures: Intellectual frameworks and research challenges. (2013).

[37] Carolyn Ellis and Art Bochner. 2000. Autoethnography, personal narrative, reflexivity: Researcher as subject. (2000).

[38] Virginia Eubanks. 2018. *Automating inequality: How high-tech tools profile, police, and punish the poor.* St. Martin's Press.

[39] Megan Alyssa Fletcher. 2018. We to Me: An Autoethnographic Discovery of Self, In and Out of Domestic Abuse. *Women's Studies in Communication* 41, 1 (Jan. 2018), 42–59. https://doi.org/10.1080/07491409.2017.1423526 Publisher: Routledge _eprint: https://doi.org/10.1080/07491409.2017.1423526.

[40] Keolu Fox. 2020. The illusion of inclusion—the "All of Us" research program and indigenous peoples' DNA. *New England Journal of Medicine* 383, 5 (2020), 411–413.

[41] 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. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 1–24. https://doi.org/10.1145/3359304

[42] 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*. 1–13.

[43] Brett M Frischmann, Michael J Madison, and Katherine Jo Strandburg. 2014. *Governing knowledge commons.* Oxford University Press.

[44] Lisa Frohmann. 2005. The Framing Safety Project: Photographs and Narratives by Battered Women. *Violence Against Women* 11, 11 (Nov. 2005), 1396–1419. https://doi.org/10.1177/1077801205280271 Publisher: SAGE Publications Inc.

[45] Candace C. Fuller, Austin Cosgrove, Kenneth E. Sands, Karla M. Miller, Russell E. Poland, Edward Rosen, Alfred Sorbello, Henry Francis, Robert Orr, Sarah K. Dutcher, Greg Measer, and Noelle M. Cocoros. 2021. Using inpatient electronic medical records to study influenza for pandemic preparedness. *Influenza and Other Respiratory Viruses* 16 (2021), 265 – 275.

[46] Christina Harrington, Sheena Erete, and Anne Marie Piper. 2019. Deconstructing community-based collaborative design: Towards more equitable participatory design engagements. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–25.

[47] Bridget A Harris and Delanie Woodlock. 2019. Digital Coercive Control: Insights From Two Landmark Domestic Violence Studies. *The British Journal of Criminology* 59, 3 (April 2019), 530–550. https://doi.org/10.1093/bjc/azy052

[48] Sam Havron, Diana Freed, Rahul Chatterjee, Damon McCoy, Nicola Dell, and Thomas Ristenpart. 2019. Clinical computer security for victims of intimate partner violence. In *28th USENIX Security Symposium (USENIX Security 19)*. 105–122.

[49] Tad Hirsch. 2020. Practicing without a license: Design research as psychotherapy. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–11.

[50] Michael Hockenhull and Marisa Leavitt Cohn. 2021. Speculative Data Work & Dashboards: Designing Alternative Data Visions. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–31.

[51] Thomas D Hull, Jacob Levine, Niels Bantilan, Angel N Desai, and Maimuna S Majumder. 2021. Analyzing Digital Evidence From a Telemental Health Platform to Assess Complex Psychological Responses to the COVID-19 Pandemic: Content Analysis of Text Messages. *JMIR Formative Research* 5, 2 (2021), e26190.

[52] Jane Im, Jill Dimond, Melody Berton, Una Lee, Katherine Mustelier, Mark S. Ackerman, and Eric Gilbert. 2021. Yes: Affirmative Consent as a Theoretical Framework for Understanding and Imagining Social Platforms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–18. https://doi.org/10.1145/3411764.3445778

[53] Eun Seo Jo and Timnit Gebru. 2020. Lessons from archives: Strategies for collecting sociocultural data in machine learning. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 306–316.

[54] Helena Karasti, Karen S Baker, and Eija Halkola. 2006. Enriching the notion of data curation in e-science: data managing and information infrastructuring in the long term ecological research (LTER) network. *Computer Supported Cooperative*

Work (CSCW) 15, 4 (2006), 321–358.

[55] Jane Kaye, Edgar A Whitley, David Lund, Michael Morrison, Harriet Teare, and Karen Melham. 2015. Dynamic consent: a patient interface for twenty-first century research networks. *European journal of human genetics* 23, 2 (2015), 141–146.

[56] Erin Kenneally and David Dittrich. 2012. The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research. *SSRN Electronic Journal* (2012). https://doi.org/10.2139/ssrn.2445102

[57] Michael Klompas, Jason McVetta, Ross Lazarus, Emma M. Eggleston, Gillian A Haney, Benjamin Kruskal, W Katherine Yih, Patricia Daly, Paul Oppedisano, Brianne M Beagan, Michael Lee, Chaim Kirby, Dawn M Heisey-Grove, Alfred DeMaria, and Richard Platt. 2012. Integrating clinical practice and public health surveillance using electronic medical record systems. *American journal of public health* 102 Suppl 3 (2012), S325–32.

[58] Kailee Kodama Muscente. 2020. The History of the Belmont Report | IRB Blog | Institutional Review Board | Teachers College, Columbia University. https://www.tc.columbia.edu/institutional-review-board/irb-blog/the-history-of-the-belmont-report/

[59] Crystal Lee, Tanya Yang, Gabrielle D Inchoco, Graham M Jones, and Arvind Satyanarayan. 2021. Viral visualizations: How coronavirus skeptics use orthodox data practices to promote unorthodox science online. In *Proceedings of the 2021 CHI conference on human factors in computing systems.* 1–18.

[60] Karen Levy and Bruce Schneier. 2020. Privacy threats in intimate relationships. *Journal of Cybersecurity* 6, 1 (Jan. 2020), tyaa006. https://doi.org/10.1093/cybsec/tyaa006

[61] Matteo Malgaroli, Thomas Derrick Hull, Niels Bantilan, Bonnie Ray, and Naomi Simon. 2020. Suicide Risk Automated Detection Using Computational Linguistic Markers From Patients' Communication With Therapists. *Biological Psychiatry* 87, 9 (2020), S444.

[62] Thomas Markussen and Eva Knutz. 2013. The poetics of design fiction. In *Proceedings of the 6th International Conference on Designing Pleasurable Products and Interfaces (DPPI '13).* Association for Computing Machinery, New York, NY, USA, 231–240. https://doi.org/10.1145/2513506.2513531

[63] Bear Montique, Peter Kelley, and Sumanta Roy. 2020. *London Domestic Homicide Review (DHR) Case Analysis and Review of Local Authorities DHR Process.* Technical Report. Standing Together, 17th September 2020. 138 pages. https://www.standingtogether.org.uk/s/Standing-Together-London-DHR-Review-Report.pdf

[64] Niamh Moore. 2006. The contexts of context: Broadening perspectives in the (re) use of qualitative data. *Methodological Innovations Online* 1, 2 (2006), 21–32.

[65] Gaia Mosconi, Qinyu Li, Dave Randall, Helena Karasti, Peter Tolmie, Jana Barutzky, Matthias Korn, and Volkmar Pipek. 2019. Three gaps in opening science. *Computer Supported Cooperative Work (CSCW)* 28, 3 (2019), 749–789.

[66] Helen Nissenbaum. 2004. Privacy as Contextual Integrity. *Washington Law Review* 79, 1 (Feb. 2004), 119. https://digitalcommons.law.uw.edu/wlr/vol79/iss1/10

[67] Safiya Umoja Noble. 2018. Algorithms of oppression. In *Algorithms of Oppression.* New York University Press.

[68] OpenNotes. 2023. Federal Rules Mandating Open Notes. https://www.opennotes.org/onc-federal-rule/

[69] Gwen Ottinger. 2010. Buckets of resistance: Standards and the effectiveness of citizen science. *Science, Technology, & Human Values* 35, 2 (2010), 244–270.

[70] Mauro Paulino, Paulo Barbosa Marques, Miguel Oliveira Rodrigues, and Stephen J. Morewitz. 2022. Chapter 11 - Contributions of psychological science to enhancing law enforcement agencies' response to intimate partner violence. In *Police Psychology*, Paulo Barbosa Marques and Mauro Paulino (Eds.). Academic Press, 205–241. https://doi.org/10.1016/B978-0-12-816544-7.00011-5

[71] Aare Puussaar, Ian G. Johnson, Kyle Montague, Philip James, and Peter Wright. 2018. Making Open Data Work for Civic Advocacy. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (Nov. 2018), 143:1–143:20. https://doi.org/10.1145/3274412

[72] David Ribes and Charlotte P Lee. 2010. Sociotechnical studies of cyberinfrastructure and e-research: Current themes and future trajectories. *Computer Supported Cooperative Work (CSCW)* 19, 3 (2010), 231–244.

[73] Phoebe Sengers, Kirsten Boehner, Shay David, and Joseph 'Jofish' Kaye. 2005. Reflective design. In *Proceedings of the 4th decennial conference on Critical computing: between sense and sensibility (CC '05).* Association for Computing Machinery, New York, NY, USA, 49–58. https://doi.org/10.1145/1094562.1094569

[74] Phoebe Sengers, John McCarthy, and Paul Dourish. 2006. Reflective HCI: Articulating an Agenda for Critical Practice. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems* (Montréal, Québec, Canada) *(CHI EA '06).* Association for Computing Machinery, New York, NY, USA, 1683–1686. https://doi.org/10.1145/1125451.1125762

[75] Divya Shanmugam and Emma Pierson. 2021. Quantifying Inequality in Underreported Medical Conditions. *arXiv preprint arXiv:2110.04133* (2021).

[76] Katie Shilton, Emanuel Moss, Sarah A. Gilbert, Matthew J. Bietz, Casey Fiesler, Jacob Metcalf, Jessica Vitak, and Michael Zimmer. 2021. Excavating awareness and power in data science: A manifesto for trustworthy pervasive data research. *Big Data & Society* 8, 2 (July 2021). https://doi.org/10.1177/20539517211040759

[77] Ryan C. Shorey, Tara L. Cornelius, and Kathryn M. Bell. 2011. Reactions to participating in dating violence research: are our questions distressing participants? *Journal of Interpersonal Violence* 26, 14 (Sept. 2011), 2890–2907. https://doi.org/10.1177/0886260510390956

[78] Dilruba Showkat and Eric P. S. Baumer. 2022. "It's Like the Value System in the Loop": Domain Experts' Values Expectations for NLP Automation. In *Designing Interactive Systems Conference (DIS '22)*. Association for Computing Machinery, New York, NY, USA, 100–122. https://doi.org/10.1145/3532106.3533483

[79] Rebecca Skloot. 2017. *The immortal life of Henrietta Lacks*. Broadway Paperbacks.

[80] Logan Stapleton, Min Hun Lee, Diana Qing, Marya Wright, Alexandra Chouldechova, Ken Holstein, Zhiwei Steven Wu, and Haiyi Zhu. 2022. Imagining new futures beyond predictive systems in child welfare: A qualitative study with impacted stakeholders. In *2022 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Seoul Republic of Korea, 1162–1177. https://doi.org/10.1145/3531146.3533177

[81] Evan Stark. 2013. Coercive control. *Violence against women: Current theory and practice in domestic abuse, sexual violence and exploitation* (2013), 17–33.

[82] Evan Stark and Marianne Hester. 2019. Coercive control: Update and review. *Violence against women* 25, 1 (2019), 81–104.

[83] Jina Suh, Eric Horvitz, Ryen W White, and Tim Althoff. 2021. Population-scale study of human needs during the covid-19 pandemic: Analysis and implications. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. 4–12.

[84] Surviving Economic Abuse. 2022. Domestic Abuse Act. https://survivingeconomicabuse.org/what-we-do/projects-and-policy/domestic-abuse-act/

[85] Laura K Thompson, Margaret M Sugg, and Jennifer R Runkle. 2018. Adolescents in crisis: A geographic exploration of help-seeking behavior using data from Crisis Text Line. *Social science & medicine* 215 (2018), 69–79.

[86] The National Network to End Domestic Violence. 2022. Reflections on 2022 NNEDV's Year in Review. https://nnedv.org/latest_update/reflections-on-2022-nnedvs-year-in-review/

[87] 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, 1893–1909. https://www.usenix.org/conference/usenixsecurity20/presentation/tseng

[88] Emily Tseng, Diana Freed, Kristen Engel, Thomas Ristenpart, and Nicola Dell. 2021. A Digital Safety Dilemma: Analysis of Computer-Mediated Computer Security Interventions for Intimate Partner Violence During COVID-19. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3411764.3445589

[89] Emily Tseng, Fabian Okeke, Madeline Sterling, and Nicola Dell. 2020. "We can learn. Why not?" Designing Technologies to Engender Equity for Home Health Aides. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

[90] 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. https://doi.org/10.1145/3491102.3502038

[91] Effy Vayena and Alessandro Blasimme. 2017. Biomedical Big Data: New Models of Control Over Access, Use and Governance. *Journal of Bioethical Inquiry* 14, 4 (Dec. 2017), 501–513. https://doi.org/10.1007/s11673-017-9809-6

[92] Janet Vertesi, Jofish Kaye, Samantha N Jarosewski, Vera D Khovanskaya, and Jenna Song. 2016. Data Narratives: Uncovering tensions in personal data management. In *Proceedings of the 19th ACM conference on Computer-Supported cooperative work & social computing*. 478–490.

[93] Nicole Westmarland and Hannah Bows. 2018. *Researching Gender, Violence and Abuse: Theory, Methods, Action*. Routledge, London. https://doi.org/10.4324/9781315630618

[94] John Wilbanks. 2018. Design issues in e-consent. *Journal of Law, Medicine & Ethics* 46, 1 (2018), 110–118.

[95] Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* 3, 1 (2016), 1–9.

[96] Richmond Y. Wong, Vera Khovanskaya, Sarah E. Fox, Nick Merrill, and Phoebe Sengers. 2020. Infrastructural Speculations: Tactics for Designing and Interrogating Lifeworlds. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376515

[97] Richmond Y. Wong, Deirdre K. Mulligan, Ellen Van Wyk, James Pierce, and John Chuang. 2017. Eliciting Values Reflections by Engaging Privacy Futures Using Design Workbooks. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (Dec. 2017), 111:1–111:26. https://doi.org/10.1145/3134746

[98] Jonathan Zong and J. Nathan Matias. 2022. Bartleby: Procedural and Substantive Ethics in the Design of Research Ethics Systems. *Social Media + Society* 8, 1 (Jan. 2022), 20563051221077021. https://doi.org/10.1177/20563051221077021 Publisher: SAGE Publications Ltd.