Scaling Research with Data Subject Rights: Understanding Stakeholders' Perspectives

ANONYMOUS AUTHOR(S)

Using data subject rights as a research method can help monitor the activities of data controllers, thereby promoting fairness, accountability, and transparency in how individuals' personal data is handled. Having citizens exercise their rights through research studies can also empower them to protect their personal data from misuse. Such research studies, however, encounter complexities or limitations which may curb their scale. It is known that some of these complexities arise from the data subjects, who may lack the expertise to fully participate. What has been less studied is the role of the data controller, who is responsible for processing subject right requests. In this work, we aim to tackle the challenges of undertaking research with data subject rights by incorporating insights from all key stakeholders: data subjects, researchers, and Data Protection Officers (DPOs). We propose a novel citizen science framework and demonstrate and discuss this with the three stakeholder groups. Thematic analysis of semi-structured interviews indicates concerns around trust and transparency, the burden and workload on stakeholders, and data protection. Our research reveals the unexplored perspectives of DPOs and other stakeholders regarding the research and implementation of data subject rights. In addition to presenting our proposed framework, we provide actionable recommendations for researchers working in this space. Furthermore, we emphasise the importance of upholding data protection principles to enhance citizens' trust in researchers, as well as the need for further empirical research to better understand the trustworthiness of researchers. Finally, we outline some challenges and opportunities for the Human-Computer Interaction (HCI) community in the evolving field of data subject rights.

CCS Concepts: • **Applied computing** → **Law**; • **Human-centered computing** → **Collaborative and social computing design and evaluation methods**; **Collaborative and social computing**; • **Security and privacy** → **Privacy protections**.

Additional Key Words and Phrases: Data protection, Data subject rights, Research methodology, Citizen science

ACM Reference Format:

Anonymous Author(s). 2025. Scaling Research with Data Subject Rights: Understanding Stakeholders' Perspectives. 1, 1 (April 2025), 22 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Data subject rights play a key role in shaping today's data ecosystems. These rights, granted to individuals (data subjects), are designed to give them greater control over their personal data, emphasising fairness, accountability, and transparency in how companies (data controllers) utilise that data. Many notable data protection regulations around the globe, such as the California Consumer Privacy Act (CCPA) [45], the European General Data Protection Regulation (GDPR) [48], and the Canadian Personal Information Protection and Electronic Documents Act (PIPEDA) [20], enshrine these rights to data subjects—individuals who can be directly or indirectly identified by the data held by data controllers. One example is the right of access (e.g. GDPR Art 15 [48, Art 15]) through which a data subject can request from the data controller-natural or legal persons that determine the purposes and means of processing personal data-a copy of their processed personal data, information about the purpose of the processing, the category of personal data processed, the retention period, and the recipients or categories of recipients with whom the personal data have been or will be shared [48]. Another data subject right, the right to data portability, allows data subjects to obtain and reuse their data across different services [48, Art 20]. Other rights—such as the right to rectification [48, Art 16] or the right not to be subject to a decision based solely on automated processing [48, Art 22] or the right to be forgotten [48, Art 17])—can

be used to challenge inaccuracies or halt the misuse of personal data. Exercising these rights generally involves three steps: (1) identifying the appropriate data controller, typically through their website; (2) sending the request, which can be done via telephone, email, or in-person communication; and (3) correspondingly working with the data controller(s) until the request is processed [6]. Data Protection Officers (DPOs) are assigned by data controllers to handle these requests and ensure compliance with data protection law.

These rights may hold different meanings for different people — activists use them to advocate for workers' rights [39]. A notable example is the Driver's Seat Cooperative, based in the US, which helps workers access work-related data to address power imbalances between drivers and platform companies [15]. Researchers also use them as a tool for data collection in research studies [7]. For example, data subject rights can facilitate audits of data held by technology companies [58]. In such context, a researcher may tackle significant societal issues by recruiting participants and encouraging them to exercise their rights, such as requesting their data from data controllers [48, Art 15 and 20] or invoking their right to be forgotten [48, Art 17] for the deletion of their personal data. The data provided, or the process itself, can then be analysed to address important societal issues. In this paper, we consider research studies that employ data subject rights as a methodology as any academic study that utilise any data subject rights for data collection; as such, we termed them as "Data Subject Rights Driven Studies".

Data subject rights present numerous opportunities for researchers in Computer-Supported Cooperative Work (CSCW), Human-Computer Interaction (HCI), and Social computing, in particular around understanding how personal data influences human experiences and algorithmic decision-making. Given that companies increasingly use personal data to make decisions about credit, insurance, healthcare, and social security. But what happens if that data is inaccurate, biased, or used unethically? Moreover, these rights can play a crucial role in understanding automated decision-making and profiling of data subjects, and can be used to hold data controllers accountable. HCI researchers are particularly well-positioned to assess the efficacy of rights-based frameworks, identify user needs, and inform policy and jurisprudence in the digital space [31]. That said, conducting research using data subject rights may be challenging due to various technical, legal, and financial complexities, as well as issues related to time and education and awareness among citizens [19, 24, 46, 55]. These factors may make it difficult to scale such studies [24].

Given the success of citizen science in expanding research efforts and addressing societal issues while educating members of the public, we propose that citizen science could help tackle the scalability challenges faced by research studies that utilise data subject rights as a methodology. Citizen science involves an open collaboration where scientists engage the public in scientific research to broaden the scope of studies and improve the collection, compilation, analysis, and interpretation of data to solve real-world problems [26, 56]. It has proven effective in generating new discoveries in various fields, including astronomy, computer science, ecology, and medicine [56]. While some studies driven by data subject rights may not require direct access to data and may instead emphasise procedural elements involved in invoking these rights, Valkenburg et al. argue that citizen science can build participants' trust in sharing their data [31].

In this study, we investigate the challenges of enacting citizen science frameworks to address issues related to scaling research methodologies driven by data subject rights. Particularly, given the confusion, complexities, and overall concerns people seem to have about these rights. We argue that establishing a standardised framework for these studies could help navigate their inherent scalability limitations and other complexities. To support this, we developed a citizen science framework for researchers conducting such studies. This framework incorporates the concept of delegating rights within the citizen science model to tackle technical and legal complexities [19, 24]. In delegation, a more knowledgeable individual, such as a researcher, takes the lead in the process of exercising data subject rights on behalf of a citizen [5]. Participants can delegate exercising their data subject rights to a researcher or opt out and

exercise their data subject rights themselves with (or without) the help of the researcher. Additionally, we introduce distributed data processing to enhance the privacy of participants. The following questions drive our research:

- **RQ1:** What are possible impediments to implementing citizen science in studies that use data subject rights?
- **RQ2:** How can delegation affect citizen scientists' participation in studies driven by data subject rights?

To investigate these questions, we first designed a citizen science framework for researchers that use data subject rights as methodology. We then created a wireframe prototype (Section 4.3) as a platform and communication channel for presenting our ideas to key stakeholders [27]. The prototype serves solely as a research tool and is not a contribution on its own (at this stage of the work) [36]. Its primary purpose was to facilitate focused discussions with stakeholders about the framework for research studies driven by data subjects' rights through citizen science [27, 36]. Our approach follows a standard method in HCI to tackle a problem (as used e.g. in Gulotta et al. [22] or Vitale et al. [53]), and allowed us to avoid committing too many resources at the outset [27].

While many of the studies that evaluated data subject rights regime highlight shortcomings in how organisations respond to these requests—ranging from non-compliance to poor communication—these critiques often lack the insights of the data controllers themselves. As a result, some degree of implicit blame is assigned to organisations for failing to meet legal or ethical expectations, without fully exploring the challenges or constraints they may face in practice. We address this gap by including perspectives from all key stakeholder groups, including DPOs. We showcased the prototype to three stakeholder groups: citizen scientists (members of the public), researchers, and DPOs. Through qualititative analysis of semi-structured interviews with nine interviewees, we make the following contributions:

(1) Empirical research utilising data subject rights has already been performed (e.g. the work of Ausloos and Dewitte [6]). We share insights on the challenges arising from these types of studies from our discussions with three key stakeholders: citizen scientists (members of the public), researchers, and DPOs. To our knowledge, this study is the first to interact with and present the perspectives of DPOs regarding data subject rights research and implementation (Section 5).
(2) We discuss the challenges involved in, and make recommendations for, designing and implementing research studies that incorporate data subject rights as a methodology for data collection (Section 6).
(3) We explore how insights from stakeholders could inform future research in this area, particularly regarding data intermediary models [19, 47]. Data Intermediaries is a term that refers to organisations that facilitate data access and exchange on behalf of individuals or for their benefit [19, 47] (Section 6.2).
(4) We propose a novel framework for conducting scalable research studies driven by data subject rights in a citizen science context. This framework could be used to obtain large data to monitor the activities of data controllers, ultimately promoting fairness, accountability, and transparency in the use of individuals' personal data. Additionally, it could serve as a tool for educational outreach, helping to inform citizens about the potential misuse of their data (Section 3). However, given the challenges associated with data subject rights, it is essential to consult with stakeholders before committing resources to implementation. Some of our interviewees expressed support for the idea. Future work could integrate insights from our research and evaluate the framework in real-life settings.

2 BACKGROUND

2.1 Data subject rights

Over 140 countries have created data protection and privacy laws, and several of these, notably the EU GDPR, create rights for individual data subjects [43]. These rights serve to empower data subjects, providing some control over their personal data, thereby promoting data protection principles such as fairness and accountability. The GDPR provides eight such rights: the right to inform, right of access, right to rectification, right to erasure, right to restriction, right to data portability, right to object, and right to not be subject to automated decision making [48]. Other data protection and privacy laws may have these rights in various forms.

While legislators intend these laws to empower data subjects, enhance privacy protection, and ensure transparency and control over the use of personal data in digital spaces [4, 46], others view these legal frameworks differently. For example, researchers use these data subject rights as a methodology for data collection in research studies [7, 24]. Activists use these rights to fight for workers' rights and demand fair processing of their data [39]. Due to different expectations of what these rights may achieve, there could be asymmetries between the law in theory and in practice [33].

Researchers play a pivotal role in the effectiveness of rights-based data protection frameworks by supporting their implementation, evaluation, and identification of user requirements [31]. Numerous studies empirically examine data subject rights, gathering evidence within and outside Europe, for or against the effectiveness of the law in practice. The implementation of these rights depends on various factors, including awareness, user perception, fear, perceived burden, data controllers' attitudes, usability issues, and cost [19, 24, 32, 46]. Taking the right to be informed for instance, only 0.96% of users (n = 315) read the entire text of privacy policies when downloading an app or installing software [42]. On the other hand, users exhibit different attitudes when dealing with genetic data; they monitor privacy policies and proactively take steps to exercise their right of access [21]. Researchers have also examined how data controllers can design responsible technologies to enable data subjects to exercise these regulatory requirements by making direct requests [41, 52]. Some researchers use these rights to collect data to improve accountability, for example using the right to data portability to externally audit data held by data controllers of pervasive systems [58].

2.2 Scaling research with data subject rights

Mahieu et al. [33] and Giannopoulou et al. [19] both propose collective use of data subject rights to shift power imbalances in favour of citizens. In practice, however, studies using data subject rights face scalability limitations [24]. Furthermore, conducting such studies is challenging; they typically involve requests submitted only by experts, and most publications concentrate on the right of access [24]. To improve scalability, both in terms of the number of rights used and the number of data subjects involved in studies, researchers have argued for a standardised framework that encompasses all data subject rights and includes non-experts in research driven by these rights [40, 41, 57].

To comply with the right of access, many data controllers (e.g. Google, Meta) offer data download packages that may be simpler to use than exercising rights [9]. Such downloads could expand the scope of research studies by facilitating participation from non-experts. However, these data downloads are limited to studies that rely on the right of access or data portability, and not all companies provide these downloads. Data controllers are not obligated to offer data download options, and even when they are available, these tools may not include certain categories of data, such as CCTV footage , due to the challenges of separating third-party data [44]. Additionally, other regulations, like the recent Digital Services Act (DSA) provide avenues for researchers to access large datasets through data disclosure orders [49, Art 40]. These orders require online service providers to share information with specific categories of persons, including

vetted researchers [16]. The DSA, however, is applicable only within the EU and again does not assist in research that involves rights outside of the right of access or data portability. Furthermore, by obtaining data directly from service providers, researchers may not engage with data subjects. By leveraging data subject rights, researchers can ensure that data collection is conducted with the explicit consent and awareness of the individuals involved.

Another barrier to scalability in studies that use data subject rights is participant knowledge — participants may lack appropriate legal and technical skills [24]. This may manifest itself when exercising rights, but also when data controllers use different strategies to refuse data subject requests [24]. One mechanism for helping non-expert citizens is to delegate the exercising of rights to an expert third-party [19, 28, 29]. Delegation has been shown to be an acceptable way of helping the public to participating in research studies [5], although a minority of participants in one study expressed concerns about privacy and trust [25].

Our proposed framework not only provides standardised mechanisms for exercising and interpreting rights that scale across laws in different jurisdictions, but also employs delegation to help non-experts participate in studies, and provides controls and data sanitisation to mitigate privacy concerns.

2.3 Citizen science

Citizen science is a type of open collaboration where scientists involve the public in scientific research to expand the scope of studies and enhance the ability to collect, compile, and analyse data, and interpret results to address real-world problems [11, 26]. Depending on the project, the public may voluntarily participate in all stages, including formulating research questions, study design, conducting experiments, collecting and analysing data, and interpreting results [18, 50]. Disciplines including astronomy, medicine and computer science have used citizen science to collect high-quality data at scale [56]. Citizen science projects also promote public understanding of science through formal or informal learning experiences, serving as educational and outreach activities to promote science [13, 54].

The Cornell Lab of Ornithology (CLO) has successfully conducted citizen science projects for over thirty years. This experience has led them to develop a methodology for creating and implementing citizen science projects [10]. This model consists of nine steps (as shown in Table 1), which we categorise to achieve five objectives: (1) Engaging the public, including amateurs, in data collection across a wide geographical area; (2) Ensuring data quality by providing training (promoting education) and resources, such as standardised protocols and data forms, to guide participants in reporting accurate data; (3) Allowing participants to have control over their data; (4) Disseminating the impact of participants' contributions to motivate further involvement; and (5) Ensuring accessibility for volunteers, regardless of their background or prior knowledge. These objectives work together to enhance the effectiveness of citizen science initiatives. Examples of projects that have successfully used this model include a platform for birdwatchers to record their sightings [37] or for participants to track bird populations at feeding sites [38]. Our framework builds upon the CLO model to provide a scalable citizen science framework for studies using data subject rights as a methodology.

3 DATA SUBJECT RIGHTS-CITIZEN SCIENCE FRAMEWORK

3.1 Framework

Our work establishes a framework to conduct studies driven by data subject rights at a greater scale by utilising citizen science. To achieve this, we must address two key issues that CLO's model does not address for research regarding data subject rights. First, unlike traditional citizen science projects, data for these studies is sourced from interaction with data controllers, and citizens may lack the legal and technical capacity to exercise their rights [24, 33]. To overcome this

Table 1. **The revised CLO model for developing and implementing a citizen science project. Alterations in bold text.**

	Data subject rights' model for developing and implementing a citizen science project
1	Decide on a scientific question.
2	Team formation - scientist/educator/technologist/evaluator.
3	**Data protection impact assessment**, develop, test, and refine protocols, data forms, and educational support materials.
4	Participant recruitment.
5	Participant training.
6	**Data subject requests** (a) **Delegate data subject rights** (b) **Exercise data subject rights** (c) **Data donation**
7	Accept, edit, analyse, and display data.
8	Data analysis and interpretation.
9	Results dissemination.
10	Evaluation.

challenge, we incorporate the concept of delegation. Second, the use of personal data necessitates a strong emphasis on protecting the privacy of participants. To ensure this, we ask researchers to conduct a Data Protection Impact Assessment (DPIA) during the initial phase of the research design. While the DPIA is optional for some projects, it is intended to help organisations identify and minimise privacy risks associated with data processing activities, as outlined in Article 35 of the GDPR [48]. This step is particularly important for projects that may pose a high risk to the rights and freedoms of data subjects. Conducting the DPIA ensures data protection and helps reduce the risk of data breaches and misuse. Additionally, participants are asked to analyse their data themselves and report only the results, which means that researchers do not have direct access to their data. This approach not only safeguards participants' privacy but also reduces researchers' workload while educating the participants in the process.

Table 1 outlines our enhanced CLO model. The first three steps are preliminary and are conducted by the researcher. These involve: (1) formulating research questions; (2) team formation; and (3) developing protocols, data forms, and educational materials. We also incorporated the (optional) DPIA into step three. The next steps, four and five, involve participant recruitment and training. Our framework adds steps related to data subject requests, which can be made either through delegation or directly by the participants themselves. Given the complexities surrounding data subject rights, we aim to simplify the process to engage a broader audience. In this context, participants are assisted in exercising their data subject rights. This approach provides support for individuals unable or unwilling to navigate these processes on their own, while still allowing those who are capable and willing to exercise their rights independently. Importantly, delegates (researchers) will not have access to participants' data. In seven, participants explore their data and report their findings to address the research questions. Allowing participants to manipulate their data not only helps protect their privacy, but also promotes educational outreach, which is one of the primary objectives of citizen science projects [10]. Participants who wish to donate their data for research purposes may do so, with the participants removing any identifiable elements so only necessary data is donated. In eight, the researcher aggregates responses from participants and interprets the data. Finally, nine and ten involve result dissemination and project evaluationrespectively.

4 METHODOLOGY

To address our research questions, we interviewed three groups of stakeholders: DPOs, researchers, and citizen scientist. To facilitate discussions with these stakeholders, we created a wireframe prototype to discuss the idea of a citizen science project adopting our framework using a research scenario [27], following the approach used in previous studies [22, 53].

4.1 Ethical considerations statement

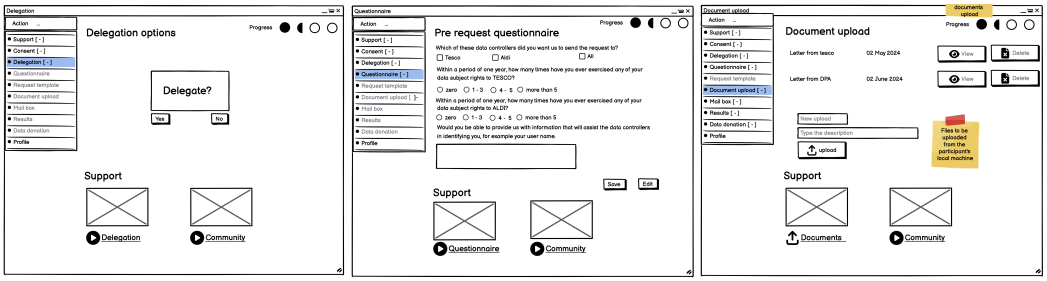
For both the prototype development and stakeholder interviews, we identified two primary ethical considerations. First, non-expert interviewees (members of the public) might not have a clear understanding of the study. We obtained informed consent. Interviewees had an opportunity to ask questions and have them answered satisfactorily. Interviewees can withdraw at any stage without giving reason(s). Second, interviewees may not want to be identified, especially DPOs. Interviewees' names (or institutions) were not attributed to their responses. Those not comfortable with their image being viewed were advised to turn off their cameras and possibly remove any identifiable data during the interview session. All responses were anonymised (and transcribed), transferred, and kept in a secure source control repository for analysis. Both the prototype and interview studies were reviewed and approved by our institutional ethics board.

4.2 The hypothetical research scenario

To illustrate the framework, we used a hypothetical research study that focuses on a fictional researcher (Sarah Johnson) investigating price discrimination among online customers based on their demographic characteristics and financial capability [1]. The study aims to explore the hypothesis that online advertisement platforms may tailor ads in a discriminatory manner based on users' demographics and financial status [1]. To address this research question, participants who have an online account with selected stores will be invited to exercise (or delegate the exercising of) their right of access or data portability and obtain the personal data held about them from data controllers. Participants from any country are welcome. The analysis will specifically target behavioral information, users' personally identifiable information (PII), device IDs, and web tracking data [2, 51].

4.3 Prototype creation

We created a wireframe prototype using the Balsamiq wireframing tool to mockup an online citizen science research project focused on investigating price discrimination (Section 4.2). The design of the prototype was informed by UX methods [27]. Through contextual inquiry and literature review, we identified key stakeholders, including citizen scientists, researchers, and DPOs, while recognising that Data Protection Authorities (DPA) have a passive role. We mapped out their roles in a collaborative workflow (Appendix 9). We studied prior works (e.g. the work of Asghari et al. [5]) and conducted a 40-minute online semi-structured interview with a researcher who has published a study driven by data subject rights. It is important to note that this researcher was not recruited for the stakeholder interview (Section 4.7). The interview questions focused on eliciting practical requirements for our prototype design. An example of a question asked was, "From your practical experience, could you run us through the step-by-step process of using data subject rights in academic studies?". The anticipated answers highlighted the factors we need to consider for various stakeholders. A detailed process was developed to guide delegation, data access, communication with research team, reporting results, and privacy-preserving data donation. We used sketching and scenario-based design to ideate and refine solutions, and created a low-to-mid-fidelity wireframe using Balsamiq.



(a) Delegation options (b) Requests questionnaire (c) Document upload

Fig. 1. **Participants can delegate the exercise of their rights or do it themselves.**

4.4 Wireframe Prototype

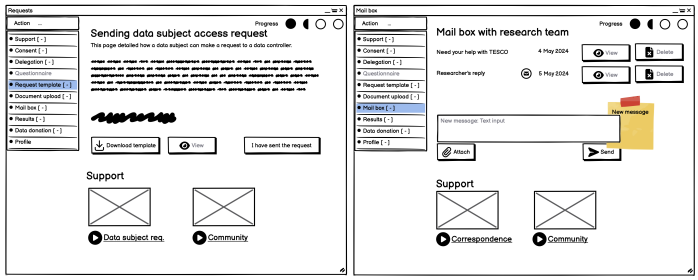
The wireframe prototype serves as a research tool to present some of our framework ideas to stakeholders and to establish a communication channel with our interviewees [27]. The prototype was designed as a fictional online interface in the form of a PDF that presented various mockup screens for the interviewees to view as if they were going through the process of being a participant in the hypothetical research study. Hartson and Pyla recommend using wireframe low-fidelity prototypes at this early stage of development to avoid committing unnecessary resources to answering research questions [27]. In the following sections, we present a walkthrough of the prototype.

4.4.1 Walkthrough of the prototype. Interviewees were shown the prototype from two perspectives, participating in the hypothetical citizen science project as both a researcher and a citizen. As a brief overview of the steps involving in being a participant of the fictional study, participants first respond to and advertisement for the citizen science research project, then if they choose to delegate, fill out a questionnaire, and requests are sent. Many steps such as responding to adverts or reporting results are similar to mainstream citizen science projects. As such, in the below paragraph we focus on describing particular features of the wireframe prototype that focus on our research questions.

Delegating the right of access: After signing up to be a participant in the fictional citizen science project, participants would need to decide to delegate their rights (Figure 1a). Upon selecting the delegation options (Figure 1a), there is a provision for recommending trusted data controllers. The data controllers will be evaluated based on ratings provided by other participants. Further details about rating data controllers will be provided in the results section. Participants who choose to delegate their rights will initially need to complete a questionnaire (Figure 1b), which guides the research team in making well-founded and appropriate requests to data controllers. Additionally, delegating participants may need to upload consent letters or documents to authenticate their identity (Figure 1c), to help the delegate certify the authenticity of the delegation requests.

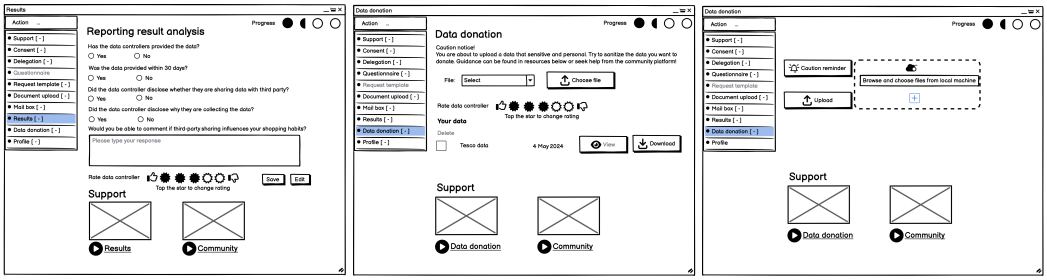
Guidance for sending data requests: Participants choose not to delegate by selecting 'No' shown in (Figure 1a), they will then receive guidance on how to make a data request themselves by providing a template that can be used for sending their data requests (Figure 2a).

Communication with researchers: After participants have opted in to take part in the citizen science project, they may want to communicate with the researcher. Participants can use a mailbox built into the interface to communicate with the research team (Figure 2b). For example, if the data controller rejects a request due to a legal provision,



(a) Sending requests (b) Mail box

Fig. 2. **Participants who opt to exercise their rights themselves are provided with a template and guidance. They can also communicate with the research team.**



(a) Results (b) Data donation A (c) Data donation B

Fig. 3. **Participants perform data analysis and submit the results, and can additionally choose to donate their data.**

participants can reach out to the researcher for further assistance. Additionally, if participants initiated the process themselves and would like the researcher to continue, they can use the mailbox to communicate their request.

Results: Regardless of whether participants delegate their rights, when participants receive data from data controllers, they will need to analyse it and submit the results to the researchers (Figure 3a). Participants are also asked to rate the data controller using a star rating system (Figures 3a & 3b). In Figure 3a, we provide example questions that could be asked for our fictional citizen science project.

Data donation: One of the final tasks for participants is to decide to donate their data for research purposes (Figures 3b & 3c). It's important to note that data donation is optional and may not be required in certain studies, particularly those that are interested in evaluating a process (or those requiring participants to exercise their right to removal [48, Art 17] or their right to restrict processing [48, Art 18]). Citizen scientists can still take part in the study without donating their data. Such data needs to be sanitised to remove identifiable elements, to try and avoid the donation of irrelevant or personally identifiable data. Participants are guided through each stage of the donation process. Before uploading the data, adequate caution regarding the removing of personally identifiable data will be provided (Figure 3b), and there is an option to review the downloaded data and delete it if necessary (Figure 3b). Previous work has developed various data donation models that can be adapted for use in this context [3, 8].

4.5 Interview sample size

Our sample size was informed by the fact that interview-based qualitative studies within HCI that focus on understanding the context and the perspectives of stakeholders often have a sample size between 5 to 15 depending on the research questions (see e.g. Designing Qualitative Research by Marshall and Rossman [34]). Additionally, we took inspiration from previous studies that utilised prototypes and interviews similar to ours; one of these studies involved 10 participants which is a similar size to ours [22]. Our emphasis was on obtaining rich insights rather than focusing solely on quantity.

4.6 Data collection

We arranged interviews with three groups of stakeholders: DPOs, researchers, and citizen scientists. All sessions were conducted remotely over Microsoft Teams and recorded using Microsoft's built-in automatic transcription. Each interview session lasted, on average, around one hour and was divided into three parts. The first part involved a topic overview learning and discussion session. We expected that DPOs and members of the general public may be unfamiliar with some terms regarding the research design, such as data donation, during the interviews. Therefore, we began by explaining our concepts, intentions, and the terms essential for understanding the topics the interview and prototype raise. For the researcher stakeholder group, it served as a refresher training.

In the second part of the sessions, we shared our screens and allowed the interviewees to take control so that we could collectively explore the design ideas presented via the wireframe prototype. Aside from the wireframe prototype we also provide interviewees with additional props to help them walk through the steps involved the fictional citizen science project. One of the authors exercised their right of access and requested their shopping data from a data controller in the UK. We modified the data and shared it as a prop with the interviewees to analyse and evaluate potential risks. We also provided them with a template for making requests, adapted from previous work [55], to explain how they could submit requests if they were comfortable doing so themselves. Additionally, we created a letter illustrating a possible response from data controllers and shared it with the interviewees. The final part of the session involved conducting the semi-structured interviews.

4.7 Interviewees & recruitment

Interviewees were recruited through personal and professional contacts of the research team. A total of nine interviewees took part: three DPOs, three researchers, and three members of the public. All three interviewees in the researcher group had conducted and published at least one peer-reviewed study using data subject rights and none of them participated in the contextual inquiry interview (Section 4.3).

4.8 Data analysis

One author conducted the interviews and recordings were automatically transcribed by Microsoft Teams. We used an inductive thematic analysis approach for the analysis, following the six steps of thematic analysis by Braun and Clarke [12]. All the researchers independently familiarised themselves with the data and generated initial codes. One researcher coded the data, taking into consideration the codes generated by the others. We had several meetings to refine the codes and share insights, after which themes were created and iteratively refined.

For reporting interview excerpts, we use identifiers comprising I (Interviewee), their group [D (DPO), R (Researcher), and C (Citizen scientist/member of the public)] and a number representing the chronological order of the interview.

5 FINDINGS

In this section, we first provide interviewees' general reactions to our work. Then, we will present the three themes that emerged from our thematic analysis. We will highlight trust and transparency as central topics interviewees discussed from multiple stakeholders' perspectives. Next, we present interviewees' views on the potential burden and workload this initiative may impose on various stakeholders. Lastly, we will discuss points related to data protection highlighted by our interviewees. This section will be divided into three parts: (1) Trust and transparency, (2) Burden and workload, and (3) Data Protection.

The majority of interviewees responded positively to our framework. Many appreciated the overall idea, particularly how incorporating delegation could simplify navigating the complexity related to citizens exercising their rights, distributed data processing, and their satisfaction with the idea of data donation. Some interviewees highlighted the importance of engaging individuals and encouraging their participation in studies employing data subject rights as a methodology, ultimately supporting the concept of citizen science. They believe that while our framework is good, involving more participation could uncover valuable insights and facilitate necessary improvements. IR02 expressed that *"The building, the system that you're doing is a very good idea. That's like a really an important part of it to have, like,...But apart from that, I think if we just more people start doing it, I think it could, that would help, like learn lessons and share lessons and kind of improve it"* (IR02). By actively engaging citizens, we can raise awareness and contribute to educational outreach efforts concerning data subject rights. We believe this empowerment would enable the public to make informed decisions about their personal data, ultimately supporting the objectives of data protection regulations.

5.1 Trust and transparency

Trust and transparency were central topics raised by interviewees. In this section, we will discuss some of the trust and transparency related perspectives explored during interviews.

Trust emerged as a significant topic, with seven of the nine interviewees identifying it as a crucial issue. IR01 said *"for this to work, we would need a trust anchor and which may be this platform that you suggest, that makes sure nothing can happen, and researchers can be relied on ...for me is one of the biggest issues that I see ...for the users, how can they trust this platform?"* (IR01). What becomes apparent from this quotation is that trust appears to be foundational, as participants need to believe that researchers will act responsibly and in their best interest. Additionally, a lack of trust can affect the willingness of individuals to delegate their data subject rights or donate their data, particularly in situations where sensitive data are involved. For example, IC02 responded to the main issues, impediments, or risks associated with delegating data subject rights in research studies by commenting: *"whether or not the subject can trust what will be done with the data? I think that would be the biggest concern"* (IC02). IR03 emphasised the importance of establishing trust with citizens participating in this kind of research *"my experience has been that citizens actually, probably once they choose to use your service or help you with research, they trust you, mostly, so they would probably be fine to let you handle really the communication"* (IR03).

By using our framework, participants place their trust in researchers to act on their behalf, whether through delegation, the use of their donated data, or the analysis of the results. IC01 expressed satisfaction with the design of our framework, stating, *"Who do you talk to? How do you ask to get access to your data? I didn't even know anything, until today, about that you could actually request that kind of data ...it would really be useful to have a delegate"* (IC01). But the consequences of researchers breaching this trust could be significant. IR03 remarked: *"I think the main risk potentially is to a kind of a breach of trust. So data subjects delegate something to the researcher and they have a certain image of what*

you're going to do as a researcher. And then if you do something different or more than they expected, that could cause distrust or and harm. It's a bit sensitive because not only might that cause distrust in that research or that researcher, but that might actually then create distrust in the system of research or the whole position of people towards researchers (IR03). IC02 suggested that improving transparency could help enhance the trust relationship with researchers, stating, *"that there should be a track record of demonstrating ways of handling people's data"* (IC02).

Transparency was another prominent topic raised by the interviewees, and could enhance trust. When asked how we can improve trust in data controllers, most interviewees cited issues related to transparency. For example, when asked this question, ID03 advocates for better transparency stating *"Better transparency. I think I find working in the profession that some people forget that at the end of these requests is a person and actually the whole issue of data rights was actually meant to improve transparency between the data subject and the controller, and I think some organisations almost start from the premise of how can we restrict this access rather than how can we provide it"* (ID03). Additionally, interviewees indicated the need for clear and open communication about the research process, data required, tasks involved, processing requests, and accountability. For example, IC02 believe that: *"If there's a way of showing people's track records for handling data. I think that it would incentivise more people to join"* (IC02). Overall, these statements suggest that transparency could build confidence in data controllers and ensure that participants are well informed, reducing concerns about misuse or mishandling of data or delegated responsibilities.

The wireframe prototype we showed interviewees had a feature for citizens to rate their experiences with data controllers. By rating citizens' experiences, we wanted to explore ways to foster trust and transparency in data subject rights research projects. Most interviewees responded positively to the idea, for example, IC02 noted, *"Getting testimonies from previous people . . . that's very useful because someone could be like, what did Laura or what did Greg say about this process?"* (IC02). Interviewees suggested various design ideas for citizens to rate their experience of data controllers, with most advocating the use of metrics. For example, when discussing design ideas to improve the ratings, ID03 mentioned metrics like: *"What the legibility of that would be, and whether I thought it was comprehensive"* (ID03). Additionally, there was a suggestion to collect quantitative data for each metric, along with qualitative feedback that captures emotional responses. *"In our experience, when we did that [data subject rights driven studies], people had like quite a lot of rich emotional experience . . . You could have this one to five, but maybe. You should also ask maybe do you want to explain it? Like maybe there should be a box where they can explain a little bit why they gave up ink and then maybe later on you could do some kind of coding analysis on that"*(IR02). ID02, however, cautioned against evaluating data controllers beyond their obligations to the data protection regulations. They suggested that rating data controllers should be based solely on whether they have met their legal obligations to avoid unfair judgments. *"I can't think that you could rate them on anything else other than have they met their obligations under the legislation . . . otherwise you could be asking people to form a judgment about a controller which was unfair"* (PDO2). We can see there is an interesting mix of opinions from different stakeholders regarding how ratings should be presented, what they should be based on regarding subjective experience or adhering to objective accountable standards. Ratings, as agreed by the interviewees, are a perfect way to show the experience of others and this would enhance transparency and by extension trust.

To sum up, trust emerged as a key topic, with participants needing to rely on researchers to handle sensitive data responsibly. Transparency in the research process and data handling was considered crucial to building confidence and encouraging participation. Many interviewees endorsed the idea of a rating system for data controllers to improve transparency and by extension trust, suggesting both quantitative and qualitative evaluation structures.

5.2 Burden and workload

This theme discusses issues of burden and workload raised by interviewees for the various stakeholders.

Interviewees identified that the proposed prototype demonstrating aspects of the framework presents a steep learning curve and a heavy workload for some stakeholders. When discussing training videos for citizens scientists taking part in the hypothetical data rights research project, ID01 referenced his wife's potential reaction: *"She would look at this and think, . . . Don't have the time. I'm not going to. I can't be bothered to watch the videos."* (ID01). The videos are intended to guide participants through key study elements, such as sanitising data for donation—removing all identifiable information and only sharing what is necessary. It was raised that this process may be too complex for non-experts, who are the likely participants. In response to the data donation steps (data sanitation) in the prototype, ID03 commented, *"My goodness, that's quite difficult"* (ID03). The added complexity for data donation could pose challenges for individuals participating in such studies, as noted by IR03: *"I am in principle not interested in how the sanitisation works. I can imagine that if that part is difficult, it might stop people in the process of doing the work"* (IR03). IC03 emphasised that for the idea to be effective in a citizen science context, this procedure *"has to be simplified"* (IC03). This underscores the challenges of engaging citizen scientists specifically in data rights research projects. Participants might inadvertently upload identifiable information, which would undermine the goal of protecting their privacy.

One interviewee stressed the need for a delegate to simplify complex processes and communication, making the process more efficient for non-experts. IC01 said, *"Because of the cumbersome nature of communicating with so many people and the time it's gonna take for them to get that data and you know how difficult some data controllers might be; it would really be useful to have a delegate who can automate or who can make that process easier and more streamlined"* (IC01). This quotation highlights how a delegate can ease the burden on citizen scientists participating in such studies.

Interviewees highlighted the challenge of overburdening data controllers' human resources. All three DPOs unanimously identified the main challenges of processing data subject requests as being complex and the significant human effort required. The data involved is often intricate and demands careful analysis. As ID01 explained, *"what is complicated is, the volume is large and trying to separate out what a person is and is not entitled to is complicated"* (ID01). To assist researchers conducting studies driven by data subject rights, ID02 offered some advice: *"You might ask individuals or to make sure that the individuals are only applying their right [data subject rights] to say if you've got 50 people in your study, then they're using 50 different controllers. Because otherwise, if you've got an organisation who has one person doing subject access requests. . . That's gonna be quite a strain on that individual from a human perspective to have 50 requests [data subject requests] that are all due on the same day or all around the same day"* (ID02).

Another concern raised about delegation is the increased verification burden it may place on data controllers. Some DPOs emphasised the extra effort required to verify the authenticity of these requests. As ID02 explained that researchers, *"Doing it [making a request on a citizens behalf] for them [citizens] I think adds unnecessary complexity to this. Particularly, adds cost to the [data controller]"* (ID02). Nevertheless, they would accept delegation requests from registered law firms, as this would simplify the verification of the legitimacy of such requests.

Researchers may need to make follow-up requests beyond the data controllers. Handling some of these requests might require the involvement of the DPAs. Some DPOs have expressed concerns that researchers may struggle with this process. For instance, ID01 noted: *"we would much rather . . . the organisation went to the ICO, and the ICO wrote to us and said this person definitely has the right to write [make a request] on this person's behalf and then we will engage with you. But the researcher will never do it"* (ID01). The pressure on researchers managing participants may lead them to

abandon follow-up communications with the DPAs. This situation could send misleading signals to policymakers and hinder other research projects as a result.

To summarise, interviewees noted that a steep learning curve and heavy workloads are potential barriers to adopting our framework. An interviewee emphasised the importance of having a delegate to help ease the burden on non-experts. The interviewees also pointed out the challenges of handling data subject requests, which are often complex and demand substantial human effort. Additionally, some DPOs expressed concerns about the increased verification burden associated with delegation requests. Researchers may struggle with follow-up requests, particularly involving DPAs, potentially leading to abandoned communications and complications in future research projects.

5.3 Data protection

In this theme, we highlight some concerns related to data protection principles. While the focus is primarily on the GDPR, we believe that the issues we highlight are also relevant to other similar legislations.

Interviewees expressed concerns about collecting additional data that might be required for citizens to delegate exercising the data subject rights to researchers, such as photographic identity for verification and the ethical use of participant's data. IR01 stated that *"my privacy concerns are that I may have to provide identifying information to the researchers"* (IR01). In today's data privacy-conscious environment, this quotation suggests that people maybe very cautious about sharing their personal information. Similarly, any potential risk to data security and ethical use of data can cause problems to citizens from participating in these types of studies. ID03 states: *"Well, the fact that you're trying to get anonymised material from the participant rather than personal data, I'd tell them to send it direct [to the data controller]. Because you don't really want to store that. There's no benefit to you [the researcher] having that information. It just causes you [the researcher] a problem, I think"* (ID03). IR01 and ID02 argue that researchers are not good at handling participants' data. IR01 states: *"We talk about a lot about identity fraud. We talk a lot about data. The data breaches . . . that is the honest researchers are not always the best people protecting your data, many of them maybe . . . have no idea how to encrypt data securely, how to manage authorisation to this data"* (IR01). Additionally, they argue that having more participants' data could lead to identifying them. In IR01's words: *"because if as a researcher you can identify your participants that is bad, then you become a very powerful data controller"* (IR01).

To address the concerns raised above, many interviewees emphasised the importance of upholding data minimisation principles when authenticating requests made by data controllers. Several interviewees proposed implementing a two-factor authentication system to authorise these requests. IC02 commented *"I feel like we may have moved past the driver's licence or, like giving a tangible piece of identification because now we're in an age where everyone's hanging around with a smartphone or a laptop"* (IC02). Interviewees suggested using one-time password (OTP) systems to verify requests, while others recommended utilising other available digital methods.

Delegating data subject rights in research studies raises concerns about legal uncertainty, particularly since the GDPR does not explicitly address it. IR01 noted that delegation might help address some of the challenges they faced during their research, such as participants struggling to understand complex legal terms; however, they are unsure about its legality. IR01 stated, *"I don't know whether the delegation was actually allowed under GDPR"* (IR01). This situation left the interviewee [a researcher] feeling uncertain, and they remained undecided about how to proceed with the idea. Similarly, ID02 opposed delegation for various reasons, among which, were concerns about the legal ambiguity associated with it, commenting, *"we're responding to them without having to go through any international data transfer agreements because the requests from an individual. And it's purely for their personal and domestic purposes.*

So we don't have to worry about the international data transfer of their personal data to them. But if we are passing it to an intermediary, we might not have a lawful basis to pass to them, particularly if they are overseas" (ID02).

Other concerns raised about delegation included the confidentiality and the security of data subjects. Some interviewees emphasised that maintaining confidentiality and security is crucial based on their experiences with data subject requests. Both ID01 and ID02 pointed out that individuals typically exercise their rights only when issues arise. ID01 stated, *"the subject access request is a sign that you are now usually in a combat relationship with the data subject. Something has gone wrong"* (ID01). This perspective may explain their cautious approach when handling these requests, as any potential data leakage could further escalate the already strained relationship. A researcher, IR03, however, expressed disagreement with the dissenting views regarding delegation, insisting that *"In most case law that I've seen, . . . controllers try to argue that delegation cannot be used. Because it's a privacy risk, for example, and I feel that mostly it's just a way for them to delay, to use a tactic, to not respond. Rather than a genuine concern over privacy"* (IR03).

All three interviewed researchers agreed that there is a lack of standardisation in how data controllers respond to data subject requests and the process is disorganised. IR02 noted, *"The main challenge is this lack of standardisation and how organisations respond. So the process was like very manual, very messy"* (IR02). IR03 argued that increased numbers of studies would be crucial to defining and streamlining the process: *"we get collectively better at it . . . we know what are the problems, we know how to solve the problems"* (IR03).

In summary, interviewees expressed concerns about data protection, particularly regarding the collection of sensitive data for verification and the ethical use of personal information. To address these issues, they suggested implementing two-factor authentication systems to enhance security. Another concern was the legal uncertainty surrounding the delegation of data subject rights under the GDPR, with some participants unsure about its legality. Confidentiality and data security were also key issues, with some opposing delegation due to potential privacy risks. Researchers highlighted a lack of standardisation in the process and emphasised the need for clearer guidelines.

6 DISCUSSION

Our results show varying levels of concern among stakeholders regarding delegation, with the prevailing opinion among DPOs being one of skepticism. Two out of the three DPOs we spoke with emphasised the importance of security and confidentiality of participants' data. This perspective is supported by existing literature. For example, research by Cagnazzo et al. and Martino et al. [14, 35] explored various models to access personal data without consent, achieving some success in their attempts. This underscores the need for heightened vigilance regarding both processing delegation and handling data subject requests. On the other hand, all the researchers interviewed believe that delegation can benefit both researchers and participants. Further studies are needed to investigate and mitigate these issues.

Our analysis reveals differing perceptions among researchers and DPOs regarding data subject requests. Some DPOs view these requests as indicators of a problem, often leading to a "combat relationship" with data subjects, as discussed in Section 5.3. DPOs often make judgments about whether data subjects are making requests due to potential issues. Some researchers, like IR02, argue that DPOs use data security concerns as a tactic to deny requests. This divergence in viewpoints could significantly hinder studies driven by data subject rights. When a key stakeholder is seen as an obstacle instead of a facilitator, both researchers and participants may hesitate to engage in these studies. This mistrust can lead to resistance to complying with legal obligations and processing requests. A perceived lack of support or overly rigid enforcement from any stakeholder could undermine the success of these studies, limiting participation and data quality while potentially increasing legal risks.

Another highlight from our interviews is that not all the stakeholder groups were aware that researchers are becoming powerful data controllers due to the personal data they manage regarding participants. When it comes to delegation, participants may need to provide additional data for researchers to properly channel the request. Guidance from the European Data Protection Board (EDPB), or similar resources, regarding the right of access to personal data should be a reference point when designing such procedures [17]. It is also important to consider that some data subjects may register with data controllers using fake identities. Most of our interviewees suggested using digital tokens to verify requests, as this approach would enhance data minimisation. While we are not there yet, we believe this could become a reality in the near future. Regarding data donations, despite warnings, individuals might still donate identifiable data. For donations to occur, there must be a level of trust between participants and researchers. Researchers must understand that trust in them and must implement measures to ensure that any data revealing patterns about participants is deleted. Below, we outline some recommendations for researchers interested in using our framework.

Our research has identified the challenges associated with implementing citizen science in this context. Future work could integrate our findings and apply our framework in real research settings. For instance, it could evaluate how the delegation process affects participation and efficiency, or explore how delegation impacts the overall user-friendliness of the process, or replicate/reproduce an existing data subject rights driven study. Other work could focus on the trustworthiness of researchers in managing these types of studies. Data subject rights have been legislated to ensure accountability and transparency among data controllers. Our findings, however, indicate that the responsibility of maintaining trust in these types of studies also falls on the researchers themselves. Previous studies have demonstrated that participants are generally willing to trust researchers that use data subject rights [25]. But other studies have shown less trust of researchers in other contexts such as microtask markets [30]. This discrepancy underscores the importance of evaluating how researchers handle participants' data in these studies.

6.1 Recommendations to researchers using data subject rights driven studies

(1) DPOs highlighted that processing data subject right requests require significant human resources (Section 5.2). Additionally, requests for delegation can lead to extra verification costs for data controllers (Section 5.2). This added burden may strain data controllers' ability to process a high volume of (or complex) requests within the legal time limits. Therefore, researchers conducting these types of studies on scale should plan for a longer time frame and aim to involve multiple data controllers rather than relying on just a few.

(2) Researchers need to recognise that the data provided by data controllers may contain information about participants (citizens) that could inadvertently cause harm. One of our citizen interviewees, IC01, expressed that they were unaware of certain types of data being stored and accessed prior to their participation in our study (Section 5.1). Additionally, DPOs pointed to the difficulty of distinguishing between what information pertains to a person and what does not (Section 5.2). This complexity can lead to human error and the potential for providing incorrect data. Therefore, researchers should implement suitable warnings and screenings for participants and take proactive measures to mitigate or eliminate any potential risks.

(3) Some requests may need the intervention of DPAs (Section 5.2). It would be beneficial to have someone with technical knowledge and an understanding of data protection legislation as part of the research team. This presence would assist in negotiations (or resolving disputes) with data controllers.

(4) Researchers must acknowledge that they are acting as data controllers and hold a position of trust (Sections 5.1 and 5.3). Any data breach could have serious consequences. Therefore, researchers should conduct a cybersecurity evaluation as part of their data protection impact assessment (DPIA) or ethics application. This evaluation

should cover the data they plan to collect and store, the retention period for that data, and the platforms they intend to use. It may also be beneficial to have IT services perform a security assessment. Researchers should avoid requesting (or asking participants to request) all data held about their participants; instead, they should ask only the specific data necessary for their studies.

6.2 Challenges and opportunities for HCI and Policy

Between March 17th, 2025, and May 12th, 2025, the UK's Department for Science, Innovation and Technology launched a call for evidence to shape future policy on data access and governance [47]. The aim of this initiative was to help design effective policies that promote trusted mechanisms for data access and sharing while addressing and mitigating potential risks associated with third-party control of personal data [47]. According to the government's statement: *"While some data intermediary models are already operating, many remain theoretical. Through engagement with businesses, research organisations and other bodies, it has been highlighted that a range of barriers exist to the development of data intermediaries in the UK and the potential consumer, citizen and economic benefits different models could bring." [47]* Our discussion align with these concerns, particularly regarding the issue of delegation. One critical issue is that the GDPR does not provide explicit guidance on how delegation should be managed, resulting in a grey area in both interpretation and implementation [19]. On the other hand, the UK's ICO has offered practical guidance [28], but there remains significant ambiguity and inconsistency in how delegation mechanisms are applied in real-world situations. Notably, the opinions of two of our DPO interviewees, whom are both UK based, conflict with the guidelines set forth by the ICO [28]. We believe increased engagement in this area—through research and the exercise of these rights—could provide valuable guidance, which can be enhanced through court rulings, regulatory guidance, and the sharing of insights and experiences.

The lack of clarity in this area presents both challenges and opportunities for the HCI community. As noted by Jakobi et al., HCI researchers have a unique set of methods and perspectives that can significantly contribute to legal and policy debates [31], particularly where digital systems intersect with social and ethical issues. One such intersection is the evolving field of data intermediaries and delegation, which urgently needs interdisciplinary input. Currently, this area of research is underdeveloped.

Our interviewees raised important and thought-provoking questions that highlight future research directions. For instance, how can we design technological solutions to verify delegation requests in ways that are secure, privacy-preserving, and user-friendly? As reliance on physical forms of identification—such as driver's licenses and other tangible credentials—becomes increasingly outdated, alternative approaches must be developed. One interviewee questioned how we can validate a delegate's identity and intent without resorting to outdated or exclusionary verification methods. These reflections emphasise the need for innovative technical, legal, and social solutions to support the responsible use of delegation in digital contexts.

7 LIMITATIONS

Our example research scenario involves relatively low-risk data (data collected by high-street retailers). Other studies may involve more sensitive data, such as a study investigating how medical companies treat people. In such cases, the subject access request might require participants to request information about syringes for diabetic use, for example. This could involve a complex procedure that we have not explored. Another limitation of our work is that we attempted to interview the DPA representatives, who are the policymakers, in three different jurisdictions, but our efforts were not fruitful. Their viewpoints may have impacted our work. Additionally, while we used fictional data in our prototype,

prior research indicates that utilising real data is far more valuable [23]. To mitigate this concern and help interviewees assess privacy risks, we did present participants with a sample of real data obtained by one of the authors (Section 4.6). Another limitation is that our focus was primarily on the GDPR; however, we believe our findings are applicable to similar legislation. Lastly, this is just one method for implementing a prototype of the framework, and there are several other approaches that could be explored.

8 CONCLUSION

Leveraging data subject rights in research studies offers researchers an ethical, effective, and participant-centered approach to collecting and analysing personal data, evaluating legal frameworks, improving the design of technological tools, and ensuring fairness, accountability, and transparency from data controllers. These studies, however, often face challenges related to scalability and complexity. To address these challenges and enhance accessibility for non-experts, we propose adopting a citizen science approach. This method aims to promote fairness, accountability, transparency, and compliance with privacy regulations while fostering trust and empowering participants. In this work, we introduce a scalable citizen science framework for researchers, which could contribute valuable insights for conducting studies that utilise data subject rights as a methodology on a larger scale. As this area of research is still in its infancy stage [7], there is a significant need to gather perspectives from stakeholders before committing resources to implementation. To achieve this, we used a UX-style method and conducted an investigation involving a variety of stakeholders, making a unique contribution by engaging with Data Protection Officers (DPOs), who are often overlooked in this area of research. Through thematic analysis, we identified critical concerns around trust and transparency, burden and workload, and data protection implications. Based on our discussions with interviewees, we also offer recommendations for designing and implementing studies driven by data subject rights, along with some challenges and opportunities for Human-Computer Interaction (HCI) and policy development.

REFERENCES

[1] Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. 2019. Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 1–30. https://doi.org/10.1145/3359301

[2] Athanasios Andreou, Giridhari Venkatadri, Oana Goga, Krishna P Gummadi, Patrick Loiseau, and Alan Mislove. 2018. Investigating ad transparency mechanisms in social media: A case study of Facebook's explanations. In *Proceedings of the Network and Distributed System Security Symposium (NDSS)*. San Diego, CA, USA. https://doi.org/10.14722/ndss.2018.23191

[3] Theo Araujo, Jef Ausloos, Wouter van Atteveldt, Felicia Loecherbach, Judith Moeller, Jakob Ohme, Damian Trilling, Bob van de Velde, Claes De Vreese, and Kasper Welbers. 2022. OSD2F: An open-source data donation framework. *Computational Communication Research* 4, 2 (Oct. 2022), 372–387. https://doi.org/10.5117/CCR2022.2.001.ARAU

[4] Article 29 Data Protection Working Party. 2017. *Guidelines on the right to data portability.* Technical Report. European Union, Directorate General Justice and Consumers, B-1049 Brussels, Belgium, Office No MO59 05/35. https://ec.europa.eu/newsroom/article29/items/611233

[5] Hadi Asghari, Thomas van Biemen, and Martijn Warnier. 2021. Amplifying Privacy: Scaling Up Transparency Research Through Delegated Access Requests. In *5th Workshop on Technology and Consumer Protection (ConPro '21)*. IEEE, Kauai, HI, USA. https://www.ieee-security.org/TC/SPW2021/ConPro/papers/asghari-conpro21.pdf

[6] J. Ausloos and P. Dewitte. 2018. Shattering one-way mirrors - data subject access rights in practice. *International Data Privacy Law* 8, 1 (March 2018), 4–28. https://doi.org/10.1093/idpl/ipy001

[7] Jef Ausloos and Michael Veale. [n. d.]. Researching with data rights. *Technology and Regulation* 2020 (Jan. [n. d.]), 136–157. https://doi.org/10.26116/TECHREG.2020.010

[8] Alex Berke, Robert Mahari, Alex Pentland, Kent Larson, and Dana Calacci. 2024. Insights from an Experiment Crowdsourcing Data from Thousands of US Amazon Users: The importance of transparency, money, and data use. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW (Nov. 2024), 1–48. https://doi.org/10.1145/3687005

[9] Laura Boeschoten, Roos Voorvaart, Ruben Van Den Goorbergh, Casper Kaandorp, and Martine De Vos. 2021. Automatic de-identification of data download packages. *Data Science* 2, 4 (Oct. 2021), 101–120. https://doi.org/10.3233/DS-210035

[10] Rick Bonney, Caren B. Cooper, Janis Dickinson, Steve Kelling, Tina Phillips, Kenneth V. Rosenberg, and Jennifer Shirk. 2009. Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. *BioScience* 59, 11 (12 2009), 977–984. https://doi.org/10.1525/bio.2009.59.11.9

[11] Anne Bowser, Katie Shilton, Jenny Preece, and Elizabeth Warrick. 2017. Accounting for privacy in citizen science: Ethical research in a context of openness. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. Association for Computing Machinery, New York, NY, USA, 2124–2136. https://doi.org/10.1145/2998181.2998305

[12] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (July 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[13] Dominique Brossard, Bruce Lewenstein, and Rick Bonney. 2011. Scientific knowledge and attitude change: The impact of a citizen science project. *International Journal of Science Education* 27, 9 (March 2011), 1099–1121. https://doi.org/10.1080/09500690500069483

[14] Matteo Cagnazzo, Thorsten Holz, and Norbert Pohlmann. 2019. GDPiRated – Stealing Personal Information On- and Offline. In *Proceedings of the European Symposium on Research in Computer Security (ESORICS)*. 367–386. https://doi.org/10.1007/978-3-030-29962-0_18

[15] Driver's Seat. 2024. Driver's Seat homepage. Retrieved April. 07, 2024 from https://www.driversseat.co/

[16] Laura Edelson, Inge Graef, and Filippo Lancieri. 2023. *Access to Data and Algorithms: For an Effective DMA and DSA Implementation*. Technical Report. Centre on Regulation in Europe (CERRE), Avenue Louise, 475 (box 10) 1050 Brussels, Belgium. https://cerre.eu/publications/access-to-data-and-algorithms-for-an-effective-dma-and-dsa-implementation/

[17] European Data Protection Board (EDPB). 2023. *Guidelines 01/2022 on data subject rights - Right of access*. Technical Report. European Data Protection Board (EDPB), Rue Montoyer 30, B-1000 Brussels. https://www.edpb.europa.eu/system/files/2023-04/edpb_guidelines_202201_data_subject_rights_access_v2_en.pdf

[18] Ria Follett and Vladimir Strezov. 2015. An analysis of citizen science based research: usage and publication patterns. *PloS one* 10, 11 (Nov. 2015), e0143687. https://doi.org/10.1371/journal.pone.0143687

[19] Alexandra Giannopoulou, Jef Ausloos, Sylvie Delacroix, and Heleen Janssen. 2022. Intermediating data rights exercises: the role of legal mandates. *International Data Privacy Law* 12, 4 (Nov. 2022), 316–331. https://doi.org/10.1093/idpl/ipac017

[20] Government of Canada. 2000. Personal Information Protection and Electronic Documents Act. Retrieved Apr. 7, 2025 from https://laws-lois.justice.gc.ca/ENG/ACTS/P-8.6/FullText.html/

[21] Sukeshini A Grandhi and Linda Plotnick. 2022. Do I Spit or Do I Pass? Perceived Privacy and Security Concerns of Direct-to-Consumer Genetic Testing. *Proceedings of the ACM on Human-Computer Interaction* 6, 19 (Jan. 2022), 1–26. https://doi.org/10.1145/3492838

[22] Rebecca Gulotta, William Odom, Jodi Forlizzi, and Haakon Faste. 2013. Digital artifacts as legacy: exploring the lifespan and value of digital data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1813–1822. https://doi.org/10.1145/2470654.2466240

[23] Rebecca Gulotta, Alex Sciuto, Aisling Kelliher, and Jodi Forlizzi. 2015. Curatorial Agents: How Systems Shape Our Understanding of Personal and Familial Digital Information. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA, 3453–3462. https://doi.org/10.1145/2702123.2702297

[24] Adamu Adamu Habu and Tristan Henderson. 2023. Data Subject Rights as a Research Methodology: A Systematic Literature Review. *Journal of Responsible Technology* 16 (Dec. 2023), 100070. https://doi.org/10.1016/j.jrt.2023.100070

[25] Adamu Adamu Habu and Tristan Henderson. 2023. Enhancing Citizen Participation Through Data Subject Right Delegation. In *Proceedings of IFIP WG 13.8 workshop on HCI for Digital Democracy and Citizen Participation (INTERACT, 2023)*. 1–6. https://doi.org/10.1007/978-3-031-61698-3_3

[26] Mordechai Muki Haklay, Daniel Dörler, Florian Heigl, Marina Manzoni, Susanne Hecker, Katrin Vohland, et al. 2021. *What is citizen science? The challenges of definition*. Springer Cham, Switzerland, Cham, 13–33. https://doi.org/10.1007/978-3-030-58278-4

[27] Rex Hartson and Pardha S Pyla. 2018. *The UX book: Agile UX design for a quality user experience* (2nd. ed.). Morgan Kaufmann, Cambridge, MA USA.

[28] Information Commissioner's Office. [n. d.]. How do we recognise a subject access request (SAR)? Retrieved Nov. 07, 2024 from https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/individual-rights/right-of-access/how-do-we-recognise-a-subject-access-request-sar/

[29] Ada Lovelace Institute. 2021. Exploring legal mechanisms for data stewardship. Retrieved Nov. 07, 2024 from https://www.adalovelaceinstitute.org/report/legal-mechanisms-data-stewardship/

[30] Jason Jacques and Per Ola Kristensson. 2013. Crowdsourcing a hit: measuring workers' pre-task interactions on microtask markets. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*. 86–93. https://doi.org/10.1609/hcomp.v1i1.13085

[31] Timo Jakobi and Maximilian von Grafenstein. 2023. *What HCI Can Do for (Data Protection) Law—Beyond Design*. Springer International Publishing, Cham, 115–136. https://doi.org/10.1007/978-3-031-28643-8_6

[32] Yujin Kwon, Ella Corren, Gonzalo Munilla Garrido, Chris Hoofnagle, and Dawn Song. 2023. SoK: The Gap Between Data Rights Ideals and Reality. https://doi.org/10.48550/arXiv.2312.01511 arXiv:2312.01511 [cs.CY]

[33] René L.P. Mahieu, Hadi Asghari, and Michel van Eeten. 2018. Collectively exercising the right of access: Individual effort, societal effect. *Internet Policy Review* 7, 3 (July 2018). https://doi.org/10.14763/2018.3.927

[34] Catherine Marshall and Gretchen B. Rossman. 2014. *Designing qualitative research* (6th ed.). Sage publications, 2455 Teller Rd, Newbury Park, CA 91320, US.

[35] Mariano D. Martino, Pieter Robyns, Winnie Weyts, Peter Quax, Wim Lamotte, and Ken Andries. 2019. Personal information leakage by abusing the GDPR "right of access"'. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. USENIX Association, 371–386. https://www.usenix.

org/system/files/soups2019-di_martino.pdf

[36] William Odom, Ron Wakkary, Youn-kyung Lim, Audrey Desjardins, Bart Hengeveld, and Richard Banks. 2016. From Research Prototype to Research Product. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 2549 – 2561. https://doi.org/10.1145/2858036.2858447

[37] Cornell Lab of Ornithology. 2024. eBird. Retrieved Nov. 08, 2024 from https://ebird.org/

[38] Cornell Lab of Ornithology. 2024. Project FeederWatch. Retrieved Nov. 08, 2024 from https://feederwatch.org/

[39] Open Society Foundations. 2019. Q and A: Fighting for Worker's Right to Data. Retrieved Nov. 24, 2022 from https://www.opensocietyfoundations.org/voices/q-and-a-fighting-for-workers-right-to-data

[40] Dominik Pins, Timo Jakobi, Alexander Boden, Fatemeh Alizadeh, and Volker Wulf. 2021. Alexa, we need to talk: A data literacy approach on voice assistants. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference*. Association for Computing Machinery, New York, NY, USA, 495–507. https://doi.org/10.1145/3461778.3462001

[41] Dominik Pins, Timo Jakobi, Gunnar Stevens, Fatemeh Alizadeh, and Jana Krüger. 2022. Finding, getting and understanding: the user journey for the GDPR'S right to access. *Behaviour & Information Technology* 41, 10 (May 2022), 2174–2200. https://doi.org/10.1080/0144929X.2022.2074894

[42] Wanda Presthus and Hanne Sørum. 2019. Consumer perspectives on information privacy following the implementation of the GDPR. *International Journal of Information Systems and Project Management* 7, 3 (May 2019), 19–34. https://doi.org/10.12821/ijispm070302

[43] Daniel J. Solove. 2023. The Limitations of Privacy Rights. *Notre Dame Law Review* 98, 3 (Mar 2023), 975–1036. https://heinonline.org/HOL/P?h=hein.journals/tndl98&i=1015

[44] Keith Spiller. 2016. Experiences of accessing CCTV data: The urban topologies of subject access requests. *Urban Studies* 53, 13 (Oct. 2016), 2885–2900. https://doi.org/10.1177/0042098015597640

[45] The People of the State of California. 2018. The California Consumer Privacy Act of 2018. *California Legislative Information* AB-375 (28 June 2018), 1–24. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375

[46] Sarah Turner and Leonie Maria Tanczer. 2024. In principle vs in practice: User, expert and policymaker attitudes towards the right to data portability in the internet of things. *Computer Law & Security Review* 52 (April 2024), 105912. https://doi.org/10.1016/j.clsr.2023.105912

[47] UK Government's Department for Science, Innovation and Technology. 2025. Open call for evidence - Data intermediaries. Retrieved Apr. 7, 2025 from https://www.gov.uk/government/calls-for-evidence/data-intermediaries/

[48] European Union. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union* L119 (4 May 2016), 1–88. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC

[49] European Union. 2022. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) (Text with EEA relevance). *Official Journal of the European Union* L 277/1 (19 Oct. 2022), 1–102. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R2065&qid=1670837883291

[50] U.S. General Services Administration. [n. d.]. About CitizenScience.gov. Retrieved Jan 16, 2024 from https://www.citizenscience.gov/about/

[51] Giridhari Venkatadri, Athanasios Andreou, Yabing Liu, Alan Mislove, Krishna P. Gummadi, Patrick Loiseau, and Oana Goga. 2018. Privacy Risks with Facebook's PII-Based Targeting: Auditing a Data Broker's Advertising Interface. In *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, San Francisco, CA, USA, 89–107. https://doi.org/10.1109/SP.2018.00014

[52] Sophie Veys, Daniel Serrano, Madison Stamos, Margot Herman, Nathan Reitinger, Michelle L Mazurek, and Blase Ur. 2021. Pursuing Usable and Useful Data Downloads Under GDPR/CCPA Access Rights via Co-Design. In *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*. USENIX Association, 217–242. https://www.usenix.org/system/files/soups2021-veys.pdf

[53] Francesco Vitale, Janet Chen, William Odom, and Joanna McGrenere. 2020. Data dashboard: exploring centralization and customization in personal data curation. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. Association for Computing Machinery, New York, NY, USA, 311–326. https://doi.org/10.1145/3357236.3395457

[54] Andrea Wiggins and Kevin Crowston. 2011. From conservation to crowdsourcing: A typology of citizen science. In *44th Hawaii international conference on system sciences*. IEEE, Kauai, HI, USA, 1–10. https://doi.org/10.1109/HICSS.2011.207

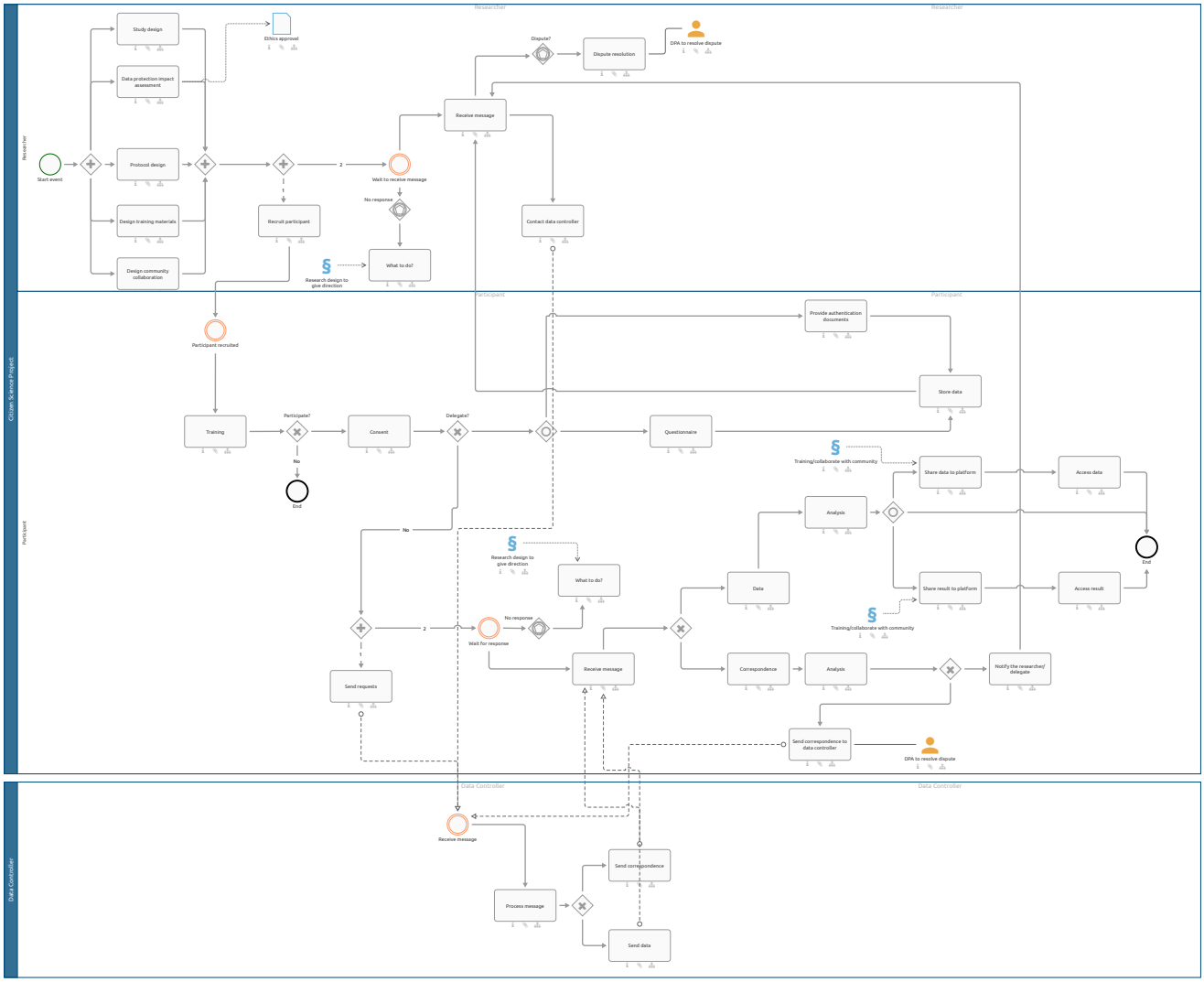
[55] Janis Wong and Tristan Henderson. 2019. The right to data portability in practice: exploring the implications of the technologically neutral GDPR. *International Data Privacy Law* 9, 3 (2019), 173–191. https://doi.org/10.1093/idpl/ipz008

[56] Jamie Woodcock, Anita Greenhill, Kate Holmes, Gary Graham, Joe Cox, Eun Young Oh, and Karen Masters. 2017. Crowdsourcing citizen science: Exploring the tensions between paid professionals and users. *Journal of Peer Production* 7, 10 (June 2017), 1099–1121. https://ora.ox.ac.uk/objects/uuid:ebe4e36c-fd05-437a-9a41-e0f5cbe0fc6a/

[57] Savvas Zannettou, Olivia Nemes-Nemeth, Oshrat Ayalon, Angelica Goetzen, Krishna P. Gummadi, Elissa M. Redmiles, and Franziska Roesner. 2024. Analyzing User Engagement with TikTok's Short Format Video Recommendations using Data Donations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA). Association for Computing Machinery, New York, NY, USA, Article 731, 16 pages. https://doi.org/10.1145/3613904.3642433

[58] Zoe Zwiebelmann and Tristan Henderson. 2021. Data Portability as a Tool for Audit. In *Proceedings of the Ubicomp Workshop on Reviewable and Auditable Pervasive Systems (WRAPS)*. Association for Computing Machinery, New York, NY, USA, 276–280. https://doi.org/10.1145/3460418.3479343

9 WORKFLOW DIAGRAM


Citizen Science Project
Researcher
Participant
Data Controller
Researcher
Researcher
Study design
Ethics approved
Data protection impact assessment
Protocol design
Design training materials
Design community collaboration
Start event
2
1
Wait to receive message
No response
What to do?
Recruit participant
Research design to give direction
Dispute?
Dispute resolution
DPA to resolve dispute
Receive message
Contact data controller
Participant
Participant
Participant recruited
Provide authentication documents
Store data
Training
Participate?
No
End
Consent
Delegate?
Questionnaire
Training/collaborate with community
Share data to platform
Access data
Analysis
End
No
Research design to give direction
What to do?
Data
Share result to platform
Access result
Training/collaborate with community
2
No response
Wait for response
Receive message
Correspondence
Analysis
Notify the researcher/ delegate
Send requests
Send correspondence to data controller
DPA to resolve dispute
Data Controller
Data Controller
Receive message
Send correspondence
Process message
Send data