*Data & Policy* (2022), 4: e3
doi:10.1017/dap.2021.40

RESEARCH ARTICLE

# Data protection for the common good: Developing a framework for a data protection-focused data commons

Janis Wong[1,*], Tristan Henderson[1] and Kirstie Ball[2]

[1]School of Computer Science, University of St Andrews, North Haugh, St Andrews KY16 9SX, United Kingdom
[2]School of Management, University of St Andrews, North Haugh, St Andrews KY16 9RJ, United Kingdom
*Corresponding author. E-mail: jccw@st-andrews.ac.uk

**Received:** 29 January 2021; **Revised:** 03 December 2021; **Accepted:** 03 December 2021

**Key words:** commons; data commons; data protection; interviews; personal data

**Abbreviations:** CPR, Common-pool resource; EU, European Union; GDPR, General Data Protection Regulation; IAD, Institutional Analysis and Development

## Abstract

In our data-driven society, personal data affecting individuals as data subjects are increasingly being collected and processed by sizeable and international companies. While data protection laws and privacy technologies attempt to limit the impact of data breaches and privacy scandals, they rely on individuals having a detailed understanding of the available recourse, resulting in the responsibilization of data protection. Existing data stewardship frameworks incorporate data-protection-by-design principles but may not include data subjects in the data protection process itself, relying on supplementary legal doctrines to better enforce data protection regulations. To better protect individual autonomy over personal data, this paper proposes a data protection-focused data commons to encourage co-creation of data protection solutions and rebalance power between data subjects and data controllers. We conduct interviews with commons experts to identify the institutional barriers to creating a commons and challenges of incorporating data protection principles into a commons, encouraging participatory innovation in data governance. We find that working with stakeholders of different backgrounds can support a commons' implementation by openly recognizing data protection limitations in laws, technologies, and policies when applied independently. We propose requirements for deploying a data protection-focused data commons by applying our findings and data protection principles such as purpose limitation and exercising data subject rights to the Institutional Analysis and Development (IAD) framework. Finally, we map the IAD framework into a commons checklist for policy-makers to accommodate co-creation and participation for all stakeholders, balancing the data protection of data subjects with opportunities for seeking value from personal data.

## Policy Significance Statement

For policy-makers, a data protection-focused data commons encourages data subject participation in the data protection process. Empowering data subjects can increase data's value when used in open and transparent ways. As our research applies participatory methodologies to collaborative data protection solutions, policy-makers can go beyond legal considerations to promote democratic decision-making within data governance. To co-create data protection solutions, we provide a checklist for policy-makers to implement a commons from the planning and public consultation process to incorporating the commons in practice. Policies for creating a commons can ease the burden of data protection authorities through preventative measures. Importantly, establishing a data protection-focused data commons encourages policy-makers to reconsider balances of power between data subjects, data controllers, and data protection stakeholders.

## 1. Introduction

Rapid technological innovation has changed how we, as individuals, interact with companies that use our personal data in our data-driven society. Data breaches and privacy scandals have frequently come to light, such as the widespread development of contact-tracing applications for invasive pandemic surveillance (Cellan-Jones, 2020), the Cambridge Analytica scandal where 50 million Facebook profiles were used to build models with the aim of influencing elections (Cadwalladr and Graham-Harrison, 2018), and the datafication of our everyday lives through the Internet of Things (Hill and Mattu, 2018). As a result, individuals are more cautious about data protection, privacy, and what information they put online (Fiesler and Hallinan, 2018; Cisco Secure 2020 Consumer Privacy Survey, 2020; Perrin, 2020; Cuthbertson, 2021; Lafontaine et al., 2021; Laziuk, 2021).

Both laws and technological solutions aim to address concerns about data breaches and privacy scandals that affect the personal data of individuals as data subjects. Data protection laws, such as the European General Data Protection Regulation (GDPR; European Union, 2016) and the California Consumer Privacy Act (California State Legislature, 2018), focus on putting responsibilities on data controllers and enforcement. As the authors are based in the United Kingdom, we focus our work on the GDPR. This regulation introduces significant changes by acknowledging the rise in international processing of big datasets and increased surveillance both by states and private companies. The GDPR clarifies the means for processing data, whereby if personal data are processed for scientific research purposes, there are safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes, or statistical purposes (Article 89), applying the principle of purpose limitation (Article 5). Data subject rights also aim to provide data subjects with the ability to better understand or prevent their data to be used, including the right of access (Article 15), the right to erasure (Article 17), and the right not to be subject to a decision based solely on automated processing (Article 22). New technologies have also attempted to give users the ability to control and recognize their sovereignty over their own data. Some tools such as Databox (Crabtree et al., 2016; a personal data management platform that collates, curates, and mediates access to an individual's personal data by verified and audited third-party applications and services) and, most prominently, Solid (Mansour et al., 2016; a decentralized peer-to-peer network of personal online data stores that allow users to have access control and storage location of their own data) prioritize creating new data infrastructures that supply online data storage entities which can be controlled by users and encourages the prevention of data-related harms as opposed to remedying harms after the fact. Other applications attempt to facilitate data reuse with privacy-by-design built in, such as The Data Transfer Project (2018; an open-source, service-to-service platform that facilitates direct portability of user data), OpenGDPR (2018; an open-source common framework that has a machine-readable specification, allowing data management in a uniform, scalable, and secure manner), and Jumbo Privacy (2019; an application that allows data subjects to backup and remove their data from platforms, and access that data locally). Such technologies help data subjects better understand the rights they may have under current regulations, as well as provide an avenue in which those rights can be acted upon.

While data protection laws and technologies attempt to address some of the potential harms caused by data breaches, they inadequately protect personal data. Current approaches to data protection rely on a high level of understanding of both the law and the resources available for individual redress. Regarding legal solutions, focusing on individual protection assumes that data subjects have working knowledge of relevant data protection laws (Mahieu et al., 2017), access to technology, and that alternatives exist to the companies they wish to break away from (Ausloos and Dewitte, 2018). Although people are more aware of their data subject rights, these are not well understood (Norris et al., 2017). Only 15% of European Union (EU) citizens indicate that they feel completely in control of their personal data (Custers et al., 2019). Evaluating location-based services, Herrmann et al. (2016) find that individuals do not necessarily know all the inferences that are made using their data and thus do not know how they are used. Importantly, individuals are unaware of, and unable to correct, false inferences, making the collection, transfer, and processing of their location data entirely opaque. Additionally, laws focusing on placing data

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

protection responsibilities on data controllers and empowering enforcement bodies assume that data controllers understand how to implement those responsibilities and that enforcement is successful (Norris et al., 2017).

While technological tools can be useful if they offer controls that limit the processing of personal data according to data subject preferences, they result in the responsibilization of data protection (Mahieu et al., 2017), where individuals have the burden of protecting their own personal data as opposed to data controllers themselves. In the case of data infrastructures, these tools may not be able to solve challenges related to the ownership of data, and it is unclear how they would meet GDPR requirements (Bolychevsky, 2021). Existing tools also frame privacy as control by placing individual onus on data protection, without supporting other GDPR principles such as data protection by design or data minimization. Tools may further assume that data subjects already have a sufficient level of understanding of the data subject rights they have by focusing on more fine-tuning privacy settings and features. They also require data subjects to trust the companies and the technological services they provide. Finally, these solutions do not offer means for collaborative data protection where information gathered from individuals could be shared among each other. This could disenfranchise data subjects from each other and prevent them from co-creating data protection solutions together through their shared experiences.

Given the limited ability for data subjects to voice their concerns and participate in the data protection process, we posit that protecting data breaches and data misuse resulting from mass data collection, processing, and sharing can be improved by actively involving data subjects in co-creation through a commons. Using commons principles and theories (E. Ostrom, 1990) and applying engagement mechanisms to innovations in our digital economy (Fung, 2015), we suggest that a commons for data protection, a "data commons," can be created to allow data subjects to collectively curate, inform, and protect each other through data sharing and the collective exercise of data protection rights. By acknowledging the limitations of data protection law and legal enforcement coupled with the desire for data-driven companies to collect and process and increasing amount of personal data, a data protection-focused data commons can help mitigate the appropriation of personal data ex post and ex ante, where data subjects are engaged throughout the process and can make the most of existing data protection regulations to co-create their own data protection solutions. In this paper, we examine how a data commons for data protection can improve data subject participation in the data protection process through collaboration and co-creation based on experiences from commons experts.

This paper is outlined as follows. First, we explore existing data stewardship frameworks and commons to better understand how they manage and govern data, identifying current solutions and why a data protection-focused data commons can support the protection data subjects' personal data through user engagement (Section 2). We then illustrate our interview methodology in identifying the challenges of building a commons and the important considerations for a commons' success (Section 3). In Section 4, we share our findings from commons experts and detail the key themes as they relate to building a commons for data protection. Based on our analysis, we then apply our findings to the Institutional Analysis and Development (IAD) framework and map the framework into an actionable checklist of considerations and recommendations for policy-makers to implement a collaborative data protection-focused data commons in practice (Section 5) as well as the next stages for deploying a commons (Section 6). Finally, we conclude that a co-created data protection-focused data commons can support more accountable data protection practices, management, and sharing for the benefit of data subjects, data controllers, and policy-makers to overcome the limitations of laws and technologies in protecting personal data.

## 2. Background

This section is split into four parts: First, we describe existing collaborative data stewardship frameworks by empirically assessing their attempt to support better data protection for data subjects through direct engagement. Then, we outline the commons and existing commons applications. Next, we identify

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

commons, urban commons, and data commons applications that are relevant to data protection. Finally, we identify theoretical applications of data protection in a data commons and illustrate our research questions to support the development of a practical framework for including data subjects in the data protection process through a data commons.

### *2.1. Data stewardship frameworks*

Data stewardship refers to the process by which "individuals or teams within data-holding organizations … are empowered to proactively initiative, facilitate, and coordinate data collaboratives toward the public interest" (Governance Lab, 2021b). Data stewards may facilitate collaboration to unlock the value of data, protect actors from harms caused by data sharing, and monitor users to ensure that their data use is appropriate and can generate data insights. New data stewardship frameworks, such as data trusts, data foundations, and data cooperatives, have been devised in order to protect data subjects as well as to involve them and other stakeholders in the co-creation of data protection solutions (Data Economy Lab, 2021b; Ada Lovelace Institute, 2021c). While these data stewardship frameworks may help mobilize data protection rights, there are significant organizational, legal, and technical differences between them. Furthermore, they also face definitional, design, and data rights-based challenges. The benefits and challenges faced by these frameworks are discussed in the following paragraphs and summarized in Table 1. A data trust is a legal structure that facilitates the storage and sharing of data through a repeatable framework of terms and mechanisms, so that independent, fiduciary stewardship of data is provided (Hardinges, 2020). Data trusts aim to increase an individual's ability to exercise their data protection rights, to empower individuals and groups in the digital environment, to proactively define terms of data use, and to support data use in ways that reflect shifting understandings of social value and changing technological capabilities (Ada Lovelace Institute, 2021b). Although data trusts refer to trusts in the legal sense, they also imply a level of trustworthy behavior between data subjects and other data trust stakeholders (O'Hara, 2019). While data trusts are promising in their ability to use trust law in order to protect data rights, it is currently unclear what powers a trustee tasked with stewarding those rights may have, and the advantages the data subjects as the trust's beneficiaries may gain (Ada Lovelace Institute, 2021b). Data trusts could in theory support responses to certain data subject rights requests, particularly through access requests, but it may be difficult to benefit from other rights such as portability and erasure to support data subjects through trusts. In the latter case, there may be tensions regarding trade secrets and intellectual property (Delacroix and Lawrence, 2019; Ada Lovelace Institute, 2021b). Moreover, the agency that data subjects may exercise within the data trust mechanism remains an open question. Efforts have encouraged the creation of "bottom-up" data trusts that aim to empower data subjects to control their data. While data subject vulnerability and their limited ability to engage with the day-to-day choices underlying data governance is acknowledged by these (Delacroix and Lawrence, 2019), many data trusts remain top-down in nature and overlook the data subject's perspective.

It is still unclear how existing fiduciary structures can fully realize their fiduciary responsibilities toward data subjects within digital spaces (McDonald, 2021; The Global Partnership of Artificial Intelligence, 2021; Data Economy Lab, 2021c). Previous pilots have attempted to clarify how data trusts can be put into practice (although without the application of trust law) by supporting the initiation and use of data trusts with a data trust life cycle (Open Data Institute, 2019). Recent projects such as the Data Trusts Initiative can help clarify how the bottom-up data trusts model can operate in practice in realizing fiduciary responsibilities (Data Trusts Initiative, 2021). Other frameworks—data foundations, data cooperatives, and data collaboratives—have also included citizen representation and engagement as an integral part of their design (Involve, 2019). There are, however, still many practical challenges in this respect (The GPAI Data Governance Working Group, 2021), particularly with questions relating to scaling and sustaining data sharing (Lewis, 2020). To address some of these challenges, data foundations have been developed as a good governance model for "responsible and sustainable nonpersonal and personal data usage, sharing, and reusage by means of independent data stewardship" (Stalla-Bourdillon et al., 2021). Data foundations rely on foundation law and view data subjects as potential beneficiaries

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

***Table 1.*** *Data trust, data foundation, data cooperative, and data collaborative stewardship models summarized by their benefits and limitations in considering data subject engagement*

| Data stewardship model | Benefits | Limitations |
|---|---|---|
| Data trust | Uses trust law as a basis for providing independent, fiduciary stewardship of data (Hardinges, 2020) with "bottom-up" inclusion of data subjects (Delacroix and Lawrence, 2019) and helps align trust and trustworthiness between them and other stakeholders (O'Hara, 2019). | There are still specific operational strategy questions that need to be answered for deployment of data trusts (The GPAI Data Governance Working Group, 2021). While there are examples of "bottom-up" data stewardship, there are no current tested examples (The Global Partnership of Artificial Intelligence, 2021) of data trusts specifically that demonstrate data subject engagement. |
| Data foundation | Provides a good governance model to minimize the risks of personal data breaches and other noncompliant data-related activities by building data usage, sharing, and reusage environments that are trustworthy by design (Stalla-Bourdillon et al., 2019). | Beneficiaries are not required within the model, and even if they are included, data subjects have limited rights in a foundation compared with a trust (Powell, 2014), with limited opportunities for direct engagement. |
| Data cooperative | Reduces the responsibilization of the data protection process for data subjects through jointly pursuing collective interests (Ada Lovelace Institute, 2021a) and can use data rights to advocate for data subjects on their behalf (P2P Foundation Wiki, 2021). | Data subjects may not be able to act independently from the group given the cooperative's group aims. Requiring contract or incorporation to establish rights and obligations could also reintroduce collaboration, engagement, and mobilization challenges (Ada Lovelace Institute, 2021a). |
| Data collaborative | Harnesses privately held data toward the public good with distinct goals to solve societal problems through collaboration between organizations from diverse sectors (Verhulst and Sangokoya, 2015). | Data collaboratives focus on exchanging data by initially mitigating risks and harms to individuals and communities (Verhulst et al., 2021), but it is unclear how individual data subjects can directly engage. |

within the model (Stalla-Bourdillon et al., 2019). However, beneficiaries are not required within the model and even if they are included, data subjects have limited rights in a foundation compared with a trust (Powell, 2014).

A data cooperative is a group that perceives itself as having collective interests, which would be better to pursue jointly than individually (Ada Lovelace Institute, 2021a). Cooperatives are "autonomous

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

associations of persons united voluntarily to meet their common economic, social, and cultural needs and aspirations through a jointly owned and democratically controlled enterprise" (International Cooperative Alliance, 2018). Data cooperatives can harness the value of data in common, where the growing real time ubiquity of digital information could help its members plan more justly and efficiently than the price mechanism in our data-driven economy (New Economics Foundation, 2018). For example, the U.S.-based Driver's Seat Cooperative is a driver-owned data cooperative that helps gig-economy workers gain access to work-related smartphone data and get insight from it with the aim to level the playing field in the gig economy (Driver's Seat, 2020).

Data cooperatives can liberate personal data through data subject access requests and can advocate for data subjects on their behalf (P2P Foundation Wiki, 2021). However, cooperatives often rely on contract or incorporation to establish rights, obligations, and governance, which could reintroduce some challenges related to collaboration and mobilization the framework was intended to limit (Ada Lovelace Institute, 2021a), where there may also be tension between reconciling individual and collective interests (Data Economy Lab, 2021a). Navigating conflict in cooperatives may be carried out through voting or other governance structures, where data cooperatives may function as fiduciaries as well (Human-Centered Artificial Intelligence, Stanford University, 2021). Similarly, data collaboratives (Governance Lab, 2021a) can harnesses privately held data toward the public good through collaboration between different sectors. Data collaboratives differ from other frameworks, such as data trusts, because the former have the distinct goal to solve societal problems through collaboration between organizations from diverse sectors (Verhulst and Sangokoya, 2015). They can support the rethinking of rights and obligations in data stewardship (Verhulst, 2021) to mitigate inequalities and data asymmetries (Young and Verhulst, 2020).

Despite many efforts to help define and clarify the legal, organizational, and technical dimensions of data trusts and other data stewardship frameworks, one of the challenges is that no broadly accepted definition of data stewardship has emerged (Stalla-Bourdillon et al., 2020). Their broad applications and widespread theoretical adoption have resulted in varied definitions and so require further disambiguation from each other in order to implement (Susha et al., 2017). Even if these frameworks are clearly defined, data trusts and data collaboratives rely on separate legal structures to facilitate the protection of personal data through the creation of a new data institution through legal means (Open Data Institute, 2021). It is acknowledged that each of these frameworks has the legal safeguarding of data subjects at their core. Nonetheless, the requirement of additional legal structures could further complicate the data protection process for both the organizations willing to adopt these frameworks as well as data subjects' ability to engage with them. Data stewardship frameworks also face several design challenges associated with the inclusion of data protection principles and data subject engagement within the framework itself. Although data protection by design may be considered (Stalla-Bourdillon et al., 2020), the frameworks may still focus more on how the data generated can be used for specific purposes as opposed to supporting data subjects' rights and agency over their personal data. While bottom-up approaches which focus on data subject agency are increasingly being considered as integral to the creation of existing data stewardship frameworks, they are not mandatory and may differ in their application. Although data subjects can be both settlors and beneficiaries within data trusts and beneficiary members in data cooperatives, individuals and groups of data subjects may still be excluded from participation in two circumstances: first, where they have not been consulted in the design of the framework, and second, where there is a lack of clarity on what a bottom-up approach entails (The Global Partnership of Artificial Intelligence, 2021). It is currently unclear how genuine and appropriate engagement mechanisms can be deployed (Ada Lovelace Institute, 2021b). Moreover, it is unclear whether or how existing data stewardship mechanisms apply participatory and action research-based solutions (Bergold and Thomas, 2012), to ensure that data subjects' preferences and perspectives are substantively taken into account as part of ongoing governance (Rabley and Keefe, 2021).

Finally, data-related rights may not be fully realized within current data stewardship frameworks. Although data stewardship frameworks benefit from not requiring extra legislative intervention that can take time to produce and is difficult to change (Ada Lovelace Institute, 2021b), the frameworks also do not

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

always interface with existing public regulatory bodies and their mechanisms, which enforce data-related rights, as part of their solution. It is also unclear how current data stewardship frameworks would support data subject recourse should there be personal data breaches. Data cooperatives often do not preserve privacy as a first priority (Ada Lovelace Institute, 2021a). While data trusts may introduce trustees and experts that are able to prevent potential data-related harms (Ada Lovelace Institute, 2021b), it is not mandatory for them to do so. Given that seeking remedies from data protection harms is not mandatory within existing data stewardship models, data subjects may be left with limited support on how to exercise data subject rights under data protection regulations.

### *2.2. The commons*

The commons, as developed by E. Ostrom, considers individual and group collective action, trust, and cooperation (E. Ostrom, 1990). The commons guards a common-pool resource (CPR), a resource system that is sufficiently large as to make it costly to exclude potential beneficiaries from obtaining benefits from its use and may be overexploited. Respecting the competitive relationships that may exist when managing a CPR, the commons depends on human activities, and CPR management follows the norms and rules of the community autonomously (E. Ostrom, 1990). The CPR enables "transparency, accountability, citizen participation, and management effectiveness" where "each stakeholder has an equal interest" (Hess, 2006). Central to governing the commons is recognizing polycentricity, a complex form of governance with multiple centers of decision-making, each of which operates with some degree of autonomy (V. Ostrom et al., 1961). Its success relies on stakeholders entering contractual and cooperative undertakings or having recourse to central mechanisms to resolve conflicts (E. Ostrom, 2010). The norms created by the commons are bottom-up, focusing on the needs and wants of the community and collectively discussing the best way to address any issues (E. Ostrom, 2012). This is illustrated by E. Ostrom's case studies of Nepalese irrigation systems, Indonesian fisheries, and Japanese mountains.

From these case studies, E. Ostrom identifies eight design principles that mark a common's success with a robust, long-enduring, CPR institution (E. Ostrom, 1990):

1. **Clearly defined boundaries**: Individuals or households who have rights to withdraw resource units from the CPR must be clearly defined, as must the boundaries of the CPR itself;
2. **Congruence between appropriation and provision rules and local conditions**: Appropriation rules restricting time, place, technology, and/or quantity of resource units are related to local cognitions and to provision rules requiring labor, material, and/or money;
3. **Collective-choice arrangement**: Most individuals affected by the operational rules can participate in modifying the operational rules;
4. **Monitoring**: Monitors, who actively audit CPR conditions and appropriate behavior, are accountable to the appropriators or are the appropriators;
5. **Graduated sanctions**: Appropriators who violate operational rules are likely to be given assessed graduated sanctions (depending on the seriousness and context of the offence), from other appropriators, by officials accountable to these appropriators, or by both;
6. **Conflict-resolution mechanisms**: Appropriators and their officials have rapid access to low-cost local arenas to resolve conflicts among appropriators or between appropriators and officials;
7. **Minimal recognition of rights to organize**: The rights of appropriators to devise their own institutions are not challenged by external governmental authorities; and
8. **For larger systems, nested enterprises for CPRs**: Appropriation, provision, monitoring, enforcement, conflict resolution, and governance activities are organized in multiple layers of nested enterprises.

As the commons on its own focuses on creating a framework to be adapted to different cases and environments, we next consider how these principles may be applied to a digital setting and data protection more specifically.

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

### *2.3. Encouraging collaboration and data subject engagement in a commons*

The commons can act as a consensus conference (Andersen and Jæger, 1999) to encourage dialogue among data subjects, experts, and policy-makers, experts and ordinary citizens, creating new knowledge together for the common good. While existing data stewardship frameworks may take their members' vested interests into consideration, the commons has a number of advantages for data subject engagement. First, a commons, through its stakeholder considerations and bottom-up norms, can directly engage data subjects in the creation and iterative improvement of the framework. Data subjects are then able to actively and continuously reflect on their individual and community preferences when it comes to managing a CPR. Second, it can advance the protection of personal data as part of democratic and participatory governance (Fung, 2015). Privacy and data protection may be addressed directly not only as legal rights, but also as part of the political, social, and cultural landscape (Dourish and Anderson, 2006). Third, the commons can offer an alternative form of data stewardship in that it applies polycentric design principles (Dourish and Anderson, 2006) and because of the commitment to these principles adopts public engagement methodologies to engage with and empower data subjects. These methodologies, which have their roots in *Human–Computer Interaction* and *Science and Technology Studies*, can increase public engagement not only with science, but also with legal, policy, and technical innovations (Wilsdon and Willis, 2004; Wilsdon et al., 2005; Stilgoe et al., 2014). Public engagement beyond the development of science, law, and policy is also necessary for establishing trust (Wynne, 2006), where the commons can support direct engagement between data subjects as well as to other stakeholders through its infrastructure as well as the application of conflict-resolution mechanisms based on E. Ostrom's design principles. Finally, the focus of these methods on worst-case scenarios—such as data breaches or privacy violations—is particularly helpful (Tironi, 2015). When addressing the likelihood of these risks occurring, the collective identification of shared goals and purpose can be enabled. Data subject agency, engagement, and empowerment may thus be garnered through the democratic expression of individual preferences toward improving individual and collective commitment toward a shared goal, while carefully juggling the interdependence between civil society and legal–political mechanisms (De Marchi, 2003).

### *2.4. Adapting the commons for transparency and accountability*

Using E. Ostrom's design principles and polycentricity as a form of governance, the commons framework has been adapted for information, data, and urban environments. These principles and frameworks can be adapted for data protection to address data-related harms by recognizing the limitations of both law and technologies, encouraging collaborative solutions for protecting personal data, and allowing data subjects to regain autonomy of their data protection process.

#### *2.4.1. Knowledge and information commons*

To address the rise of distributed, digital information, Hess and Ostrom (2007) developed the information or knowledge commons, where knowledge is the CPR. As new technologies enable the capture of information, the knowledge commons recognizes that information is no longer a free and open public good and now needs to be managed, monitored, and protected for archival sustainability and accessibility. Crucially, the commons addresses data-related governance challenges that arise due to spillovers created by the reuse of data, so increasing its value over time (Coyle, 2020). This is further exemplified when data are linked together, creating new uses and value for the same data. Socioeconomic models have also been suggested for commons-based peer production, where digital resources are created, shared, and reused in decentralized and nonhierarchical ways (Benkler et al., 2015). Without a commons, the newly generated knowledge may not be available to the original creators of the data in the first place. As a result, the knowledge commons can support data subjects in accessing the personal and social value of their data while ensuring its quality and storing it securely. More generally, commons theory has also been used to support democratic practice in digitally based societal collaborations in order to ensure diversity, define the community and the community's obligations, and build solidarity (Lee et al., 2021).

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

In assessing the feasibility of a knowledge commons, E. Ostrom's IAD framework can be used to study an institution's community, resource dynamics, and stakeholder interests. The IAD supports the creation of a commons and analyzes the dynamic situations where individuals develop new norms, rules, and physical technologies. Adopting the IAD framework's core sections on biophysical characteristics, action arena, and overall outcomes, the framework acts as a "diagnostic tool" that investigates any subject where "humans repeatedly interact within rules and norms that guide their choice of strategies and behaviors," analyzing the "dynamic situations where individuals develop new norms, new rules, and new physical technologies" (Hess and Ostrom, 2007). Institutions are defined as formal and informal rules that are understood and used by a community. Central to the IAD framework is the question "How do fallible humans come together, create communities and organizations, and make decisions and rules in order to sustain a resource or achieve a desired outcome?" Broken down into three core sections, a knowledge commons can be assessed by its resource characteristics (the biophysical–technical characteristics, community, and rules-in-use), action arena (institutional changes and the process of voluntary submitting artefacts), and overall outcomes. Specifically, for a knowledge or information commons, the IAD framework is useful, because it supports investigation into how resources are actually governed and structures the empirical inquiry to facilitate comparisons, while avoiding unwarranted assumptions related to particular theories or models (Strandburg et al., 2017). As part of the IAD framework, E. Ostrom identifies seven rules by which institutions could be analyzed (E. Ostrom, 2005):

1. **Position**: The number of possible "positions" actors in the action situation can assume (in terms of formal positions, these might be better described as job roles, while for informal positions, these might rather be social roles of some capacity);
2. **Boundary**: Characteristics participants must have in order to be able to access a particular position;
3. **Choice**: The action capacity ascribed to a particular position;
4. **Aggregation**: Any rules relating to how interactions between participants within the action situation accumulate to final outcomes (voting schemes, etc.);
5. **Information**: The types and kinds of information and information channels available to participants in their respective positions;
6. **Payoff**: The likely rewards or punishments for participating in the action situation; and
7. **Scope**: Any criteria or requirements that exist for the final outcomes from the action situation.

In advancing the practical application of the IAD framework into new use cases, the framework has been adapted to create building blocks for developing a commons. E. Ostrom adapted her design principles into key questions to created actionable means for problem-solving (E. Ostrom, 2005). Translating the IAD framework's core sections of biophysical characteristics, action arena, and overall outcomes, McGinnis transposes these questions, abstract concepts, and analytical tools to a detailed study of specific policy problems or concerns (McGinnis, 2018). McGinnis encourages users of the framework and questions to adapt them in ways that best suit the applications to the factors deemed most important for understanding the research puzzle or policy concern that serves as the focus on researchers' own work. A summary of McGinnis' steps of analysis are:

1. Decide if your primary concern is explaining a puzzle or policy analysis.
2. Summarize two to three plausible alternative explanations for why this outcome occurs, or why your preferred outcome has not been realized; express each explanation as a dynamic process.
3. Identify the focal action situation(s), the one (or a few) arena(s) of interaction that you consider to be most critical in one or more of these alternative explanations.
4. Systematically examine categories of the IAD framework to identify and highlight the most critical.
5. Follow the information flow in each of these focal action situations.
6. Locate adjacent action situations that determine the contextual categories of the focal action situation. This includes: outcomes of adjacent situations in which collective actors are constructed

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

and individual incentives shaped, rules are written and collective procedures established, norms are internalized and other community attributes are determined, goods are produced and inputs for production are extracted from resource systems (that may need replenishment), and where evaluation, learning, and feedback processes occur.

7. Compare and contrast the ways these linked and nested action situations are interrelated in the processes emphasized by each of your alternative explanations.
8. Identify the most critical steps for more detailed analysis, by isolating components of adjacent action situations that determine the context currently in place in the focal action situation(s), and that if changed would result in fundamental changes in outcomes.
9. Draw upon principles of research design or evaluative research to select cases for further analysis by whatever methods are best suited to that purpose.

When creating a knowledge commons, Strandburg et al. (2017) also mapped the IAD framework into research questions as a means to support the planning and governing process of a commons, including the interview process for gathering participants and turning those interviews into practical goals and objectives for commons governance. The knowledge commons has also been applied to privacy by considering Nissenbaum's contextual integrity (Nissenbaum, 2004) to conceptualize privacy as information flow rules-in-use constructed within a commons governance arrangement (Sanfilippo et al., 2018). We return to this in the conclusion of this paper, by discussing potential implications for policymakers of viewing privacy through an information governance lens.

An example of an information or knowledge commons is a university repository (Hess and Ostrom, 2007). Developing a university repository requires multiple layers of collective action and coordination as well as a common language and shared information and expertise. The local community, academics and researchers, can contribute to the repository, as the more it is used, the more efficient the use of resources is to the university as a public institution. Others outside that community can browse, search, read, and download the repository, further enhancing the quality of the resource by using it. By breaking down large, complex, collective action problems into action spaces through the IAD framework and using E. Ostrom's design principles for governing a commons, institutions and organizations can better meet the needs of those in the community, including how information, knowledge, and data can be used to serve the common good.

From a technological perspective, the open-source and open-software communities can also be seen as knowledge commons, where software are freely and publicly available for commercial and noncommercial uses. The software tools are also openly developed, and anyone is able to contribute. Organizations such as the Open Usage Commons (Open Usage, 2021) help project maintainers and open-source consumers have peace of mind that projects will be free and fair to use. Platforms such as Wikipedia and the Wikimedia Commons (Wikimedia, 2021) are public domain and freely licensed resource repositories that are open and can be used by anyone, anywhere, for any purpose.

E. Ostrom's design principles and the IAD framework can support a data protection-focused data commons, because they encourage active engagement of data subjects and considerations of how data can be protected through the development process while increasing its value. The analysis steps, questions, and framework encourage iterative means of creating a commons and supporting the co-creation process. The IAD framework recognizes that the expectations, possibilities, and scope of information and data can be different, as more knowledge is included within the commons. These principles are also useful in considering data protection solutions, because they recognize that there is no-one-size-fits-all fix and support more flexible and adaptable ways of achieving the commons' goals. Incorporating existing regulations and policies into the commons for data protection allows data subjects to find specific solutions to their challenges by developing a better understanding of the data protection landscape of the specific domain, collaborating with other data subjects or stakeholders to co-create individual data protection preferences, and be able to exercise their data protection rights with the support of the community that has been harmed.

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

*2.4.2. Data commons*

E. Ostrom's commons framework has been applied to data commons which guard data as a CPR. Traditionally, such data commons focus on data distribution and sharing rather than data protection (Fisher and Fortmann, 2010). Research data commons such as the Australia Research Data Commons (2020; ARDC), the Genomic Data Commons (GDC; National Cancer Institute, 2020), and the European Open Science Cloud (EOSC; European Commission, 2019) all attempt to further open-science and open-access initiatives. The ARDC is a government initiative that merges existing infrastructures to connect digital objects and increases the accessibility of research data. The National Cancer Institute also has a GDC that is used to accelerate research and discovery by sharing biomedical data using cloud-based platforms. With a research-oriented focus, the GDC does not house or distribute electronic health records or data it considers to be personally identifiable but still had safeguards against attempts to reidentify research subjects (Jensen et al., 2017). In Europe, the EOSC is a digital infrastructure set up by the European Commission for research across the EU, with the aim to simplify the funding channels between projects. The EOSC was inspired by the findable, accessible, interoperable, and reusable principles (Wilkinson et al., 2016) and aims to become a "global structure, where as a result of the right standardization, data repositories with relevant data can be used by scientists and others to benefit mankind" (European Commission, 2019). While these frameworks recognize that the information and knowledge are collectively created, their implementations are hierarchical and top-down, as they were created through structured committees, serving as a data repository platform that enables research reproducibility (Grossman et al., 2016). As a result, they may have limited input from archive participants, repository managers, or public consultation processes and do not take E. Ostrom's principles into account. Additionally, given the goals and objectives of these commons, by nature, they prioritize data sharing, data curation, and reuse, over data protection. While these data commons can be fruitful for furthering research and opening up data for reuse, they do not take into consideration the data subjects that created the data in the first place, as most data stored in these commons are not considered personally identifiable information. As a result, existing data commons alone are insufficient for protecting personal data, as they are designed without data subjects' personal data in mind.

*2.4.3. Urban commons*

While data commons may not incorporate data protection principles, some data commons frameworks applied to urban environments and urban commons have been created in an attempt for governments to take more responsibility over their citizens' personal data (European Commission, 2018). These commons are important for the development of data commons and data protection-focused commons, because they contrast other models that result in the datafication and surveillance of urban environments, such as the Alphabet Sidewalk Labs projects in Toronto (Cecco, 2020) and Portland (Coulter, 2021), both of which were scrapped due to concerns about the consolidation of data within big technology companies, lack of transparency about how public funds were to be used, and lack of public input during the development process of these smart cities. In contrast, in urban commons environments, resource management "is characteristically oriented toward use within the community, rather than exchange in the market" (Stalder, 2010). An urban commons represents resources in the city that are managed by its residents in a nonprofit-oriented and pro-social way (Dellenbaugh-Losse et al., 2020). It is a physical and digital environment that aims to better utilize an urban space for public good, formed through a participatory, collaborative process. Urban commons aim to increase the transparency of how city data are used and provide accountability should users and data subjects want their data withdrawn. For example, the European projects DECODE (European Commission, 2018) and the gE.CO Living Lab (gE.CO, 2021) both encourage citizens to be part of a collaborative process in creating communal urban environments that better represent the community. The DECODE data commons project "provides tools that put individuals in control of whether they keep their personal information private or share it for the public good" (European Commission, 2018) with the focus on city data in four different communities.

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

The project not only created an application to support user control over their data (DECODE, 2020), but also produced documents for public use on community engagement, citizen-led data governance, and smart contracts to be applied to urban environments. The outcomes from the project have been applied to local European projects such as Decidim in Barcelona to create open spaces for democratic participation for cities and organizations through free, open-source digital infrastructures (Decidim, 2021). Furthermore, the DECODE project continues to shape the EU's direction when it comes to policy-making for digital sovereignty (Bria, 2021). The gE.CO Living Lab creates "a platform for bringing together and supporting formal groups or informal communities of citizens" (gE.CO, 2021), who manage co-creation spaces and social centers created in regenerated urban voids. The Lab's aim is to foster "sharing and collaboration between citizens and establish a new partnership between public institutions and local communities, setting forth new models of governance of the urban dimension based on solidarity, inclusion, participation, economic, and environmental sustainability" (gE.CO, 2021). As cities become more digitally connected and more data are being collected from their citizens, an urban commons increasingly focuses on data both in determining how information and resources can be created and shared within a community and focusing on citizens' personal data.

### *2.5. A data commons for data protection*

More recently, organizations focusing on data governance and data stewardship have explored the use of a commons for data with applications specifically to data protection. The Ada Lovelace Institute has identified a data commons as a means to tackle data-related issues, such as consent and privacy, by mapping E. Ostrom's principles to specific GDPR principles and articles (Peppin, 2020). The focus on creating a commons for data draws attention to the sharing and dissemination of information and expertise, as it relates to data, encouraging a more open and collaborative environment. By sharing the data that are available, responsibilization can be limited, where resources are pooled for collaborative decision-making instead of individuals having to understand everything on their own. This can minimize the impact of data-related harms as a preventative method rather than a reactive one. In developing the practical basis for developing new forms of data stewardship through a commons, the Ada Lovelace Institute has also compiled a list of commons projects, mapping them to E. Ostrom's principles and creating a set of design principles for data stewardship (Ada Lovelace Institute, 2021d). More broadly looking at the value of data, the Bennett Institute and Open Data Institute have mapped E. Ostrom's principles to examples of how our data are used in a data-driven economy, highlighting the need to "provide models for sharing data that increase its use, capture positive externalities, and limit negative ones, so we can maximize the value of data to society" as well as include trustworthy institutions that together govern who can access what data "in accordance with the social and legal permissions they are given" (Bennett Institute for Public Policy and the Open Data Institute, 2020).

The data commons model for supporting data protection can be beneficial compared to existing data stewardship frameworks where data subjects and protecting personal data according to their preferences is prioritized. While data protection has been considered as part of the commons process, including data subjects and their communities is not seen as a requirement when considering how their personal data can be protected. Creating a data protection-focused data commons could help identify how much understanding and control data subjects have over their personal data and support them in choosing their data protection preferences. It can also support wider policy goals that reflect the principles, aims, and objectives as laid out by existing data protection and data-related rights through greater transparency, co-creation, and recognizing data subject agency. The consideration of supporting community norms through a commons can help ensure that the model is bottom-up in its design and iterative changes. Compared to existing data stewardship frameworks, a commons for data protection does not require the creation of a new legal framework, but rather operates within the current data infrastructures and norms used by data subjects while acknowledging the limitations of existing laws, technologies, and policies that steward data. For example, unlike the data cooperatives that require an organization to incorporate and register as a cooperative, the commons can be deployed through sociotechnical and policy means into

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

existing institutions to establish duties between stakeholders without requiring the adoption of legal stewardship requirements. This makes the commons and those who participate in it more mobile and able to react to the changes in how companies use personal data as well as data breaches. Thus, the focus on data protection as part of the data commons shifts data protection responsibilities away from the individual alone and to communities, where knowledge, expertise, and experiences can be pooled together to identify working solutions. Data subjects are able to join a specific data protection-focused data commons if they identify with the commons' aims for the protection of personal data that refers to them as individuals or a group in which they are a part of. Those who participate in a data commons should respect the community norms which they can also help create. The data commons should not only be considered as a form of personal data sharing, but rather be used as a community resource that facilitates the personal and collective aims of protecting personal data, where the sharing of personal data and data rights is not necessarily required. Anyone can leave the data commons any time they wish. Although personal data are still kept personal and private, the collaborative nature of sharing, discussion, and advising on data protection problems opens up potential options for everyone to support informed decision-making and achieving data protection preferences through a data commons. Those in the commons can then choose to act independently or as a group, whichever best suits their personal preferences. The framework can also support the remedy of potential data breaches through the exercise of data subject rights and the coordination of data rights efforts within the community. For example, a data protection-focused data commons can support the "ecology of transparency" that emphasizes the collective dimension of GDPR rights for social justice (Mahieu and Ausloos, 2021). The creation of a data protection-focused data commons can support policy goals that further the principles, aims, and objectives as laid out by data protection law through greater transparency, co-creation, and recognizing data subject agency without necessitating specific legal or technological requirements outside of community norms.

In previous work, we identify how a data protection-focused data commons can help protect data subjects from data protection harms (Wong and Henderson, 2020). A data protection-focused data commons allows individuals and groups of data subjects as stakeholders to collectively curate, inform, and protect each other through data sharing and the collective exercise of data protection rights. In a data protection-focused data commons, a data subject specifies to what extent they would like their data to be protected based on existing conflicts pre-identified within the data commons for a specific use case. An example use case would be online learning and tutorial recordings. For students (as data subjects), participating in a data protection-focused data commons allows them to better understand their school or university's policy and external organizations' guidance when it comes to collecting, processing, and sharing their personal data related to online learning, allow them to ask questions to experts, raise any questions about data protection to staff, review their consent decisions on whether to agree to tutorial recordings, and exercise their data protection rights should they wish to do so. Unlike existing data commons, the data protection-focused data commons focuses specifically on protecting data subjects' personal data with the ability to co-create and work with other data subjects, while still being able to directly exercise their data subject rights. It simplifies the data protection rights procedure by including information, instructions, and templates on how rights should be collectively exercised, giving data subjects an opportunity to engage with and shape data protection practices that govern how their personal data are protected. For example, an online learning data commons may only focus on how students' personal data are collected, used, and shared for ways to enhance learning. In contrast, an online learning data protection-focused data commons would also directly provide students with the related university policies or best practices, give students the ability to decide whether they consent to certain collection and processing of data, and support them to exercise their rights to the university's data protection officer.

### *2.6. Research questions*

As data subjects are often left out of the data protection process, they lack a meaningful voice in creating solutions that involve protecting their own personal data. Although a co-created and collaborative

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

commons has been considered for managing and protecting data, commons principles have not specifically been applied to establish a data commons that focuses on data protection and with the objective of protecting data subjects from data-related harms.

Our aim for this work was to find out more about how existing commons were created, and what the associated challenges were related to data protection, and support the implementation on a commons. To investigate those aims, we conduct interviews with commons experts to identify the challenges of building a commons and important considerations for a commons' success. From their contributions, we aim to develop a practical framework for including data subjects in the data protection process through a data commons and create a checklist to support policy-makers in implementing the data commons.

We established four research questions to explore whether using data subject rights and data protection principles to support a data protection-focused data commons is suitable both in theory and in practice:

**RQ1**: How, if at all, did interviewees work on identifying and solving data protection challenges?

**RQ2**: How can the challenges of implementing a data commons best be overcome, specifically for data protection?

**RQ3**: What do interviewees think could be done better in terms of creating a commons?

**RQ4**: Is a commons framework useful for ensuring that personal data and privacy are better protected and preserved?

## 3. Methodology

We developed our study in three phases: identifying relevant commons and key informants, writing the interview questions, and conducting the interviews.

### *3.1. Identifying relevant commons and key informants*

Urban commons and data commons applied to urban cities were identified as the most relevant to establishing a data protection-focused data commons, because they represent a commons model that considered data protection and privacy. The relevant commons identified for answering our research questions were found through conducting a literature review on recent self-described urban commons and data commons. The commons selected all used the commons to describe their work and their aims, with goals that emphasize co-creation and collaborative work with the community. As all authors reside within the jurisdiction of the GDPR, an online search was conducted to identify European commons only.

Once the commons were identified, experts were chosen based on their expertise and experience in creating and developing an urban commons or data commons, and were contacted via e-mail. To ensure that we had a fair assessment of the commons development process, when contacting experts, we made sure that they had different levels of expertise, different roles and responsibilities within commons development, and represented different communities. The size and scope of each commons project was also as varied as possible in order to better understand how similar or different commons challenges may be throughout development.

Interviews were conducted to contextualize the role of the commons from different stakeholder perspectives and provide useful information into potential challenges in the development process. Interviewees were told that this study contributes to our wider work on establishing a data protection-focused data commons to achieve better data protection for data subjects regarding the processing of their personal data in a collaborative way and allows them to co-create data protection policies with other data subjects and stakeholders, examining how information rights can be supported through a commons.

Prior to the interview, key informants were given a participant information document and a consent form for them to sign and return. Once the interview was complete, a debrief was sent to the participant with more information about their data rights and our broader research.

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

### *3.2. Writing the interview questions*

Key informant interview methods were used to design the interview, with a semi-structured format to encourage discussion around the commons. The questions aimed to answer the research questions identified in Section 2.6, augmenting what data protection lacks to explore the relevance of the creation of a data protection-focused commons and whether information rights can help with finding a solution. The interview method and questions are included as part of the Supplementary Material of this article.

In responding to the research questions identified in Section 2.6, we asked the experts those questions and supplemented them with the following questions, before engaging in further discussion:

**RQ1:**

- How did you and the project team come about identifying your project aims and what were some of the problems or challenges you considered during that process?
- What stakeholders did you interact with to solve some of these challenges?

**RQ2:**

- How did you go about solving the challenges identified?
- Were there problems or challenges during the project that you did not expect related to data?

**RQ3:**

- What do you think are/were the successes of your role in the commons?
- How was this success achieved?
- What do you think are/were the limitations of your project as a commons?
- Is there anything you would do differently?

**RQ4:**

- How do you think data and data protection can be best represented in the commons and in commoning?
- Can data subjects and participants' involvement in creating the commons support better data protection practices?
- What do you think is the ideal commons for data? Do you think it can be achieved? If so, how?

### *3.3. Conducting the interviews*

Interviews were conducted either over the phone or conferencing software, such as Skype, jit.si, or GoToMeeting, based on the interviewees' preference. All interviews were conducted by the first author between March and November 2020 and lasted up to 1 hour. All interviews were recorded with the interviewee's consent. Once each interview was completed, audio recordings were placed into the MaxQDA qualitative data analysis software for immediate transcription and pseudonymization. Once the transcription was finished, audio recordings were deleted.

## 4. Analysis

Nine experts across six commons were interviewed. The size, number of participants, and stakeholders varied across the commons, with three interviewees based in the Netherlands, two in the United Kingdom, one in Belgium, one in Germany, one in Italy, and one in Spain. Their roles and specialisms are listed in Table 2. Reference *Cx* denotes the commons they contributed to, and *Ex* denotes the expert. Role

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

characterizes the experts based on their responsibilities within the commons. Expertise describes their main contribution toward the commons.

Using MaxQDA, codes and tags were used to identify patterns for preliminary transcript analysis, identified in Figures 1 and 2. Although the interview centered around developing a commons and challenges regarding data protection, discussions around people and their interaction with others can be seen as the most prominent topic mentioned by experts. In particular, human-centered themes, such as political and financial relationships, were mentioned the most. Based on these characterizations, main themes were drawn out and expanded upon from the interviews. For our interview transcript analysis, we present our results in four sections: identifying data protection challenges, overcoming data protection challenges, improving the commons, and building a commons for data protection. We found that political and institutional barriers when it came to creating a commons were the most difficult to tackle, underlying how data and data protection are not necessarily seen as something that could be perceived as a commons. While data protection was discussed as part of the commons development process, there were limited applications to wider data protection principles such as those relating to informing data subjects about their rights and the ability to exercise those rights against data controllers. All experts identified limitations within their own area of expertise, suggesting that these limitations, whether in law, technology, or policy, need to be identified within the commons in order to find better data protection solutions. Although the decision to use a commons was to provide certain levels of control and transparency of how data were collected, used, and processed, financial restrictions limited the potential impact of the commons framework and the extent to which a commons could scale. Interviewees further mentioned

***Table 2.*** *List of interviewees representing their commons project, role within the project, and their expertise*

| Ref | Role | Expertise |
|---|---|---|
| C1E1 | Academic | Privacy and Computer Science |
| C2E1 | Technical | Privacy and Software Engineering |
| C2E2 | Governance | Public Planning and Public Policy |
| C2E3 | Policy | Commons Theory and Peer-to-Peer |
| C3E1 | Policy | Technology and Public Research |
| C3E2 | Academic | Privacy, Law, and Information Science |
| C4E1 | Policy | Third Sector and Community Engagement |
| C5E1 | Policy | Community Development Planning and Public Research |
| C6E1 | Research | Commons Policy and Community Engagement |

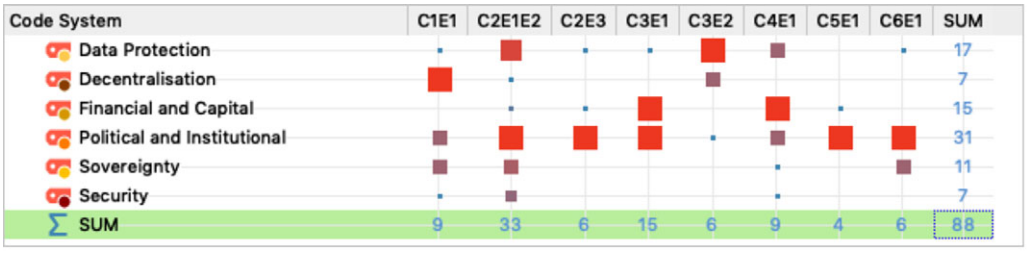


***Figure 1.*** *Code matrix created from interview transcripts with all experts. Manually coded themes related to identifying problems and challenges were tagged, and their frequencies are visualized based on how often they were discussed by interviewees. The most prominent challenges are those related to politics and institutions (31), followed by data protection (17), and financial and capital (15) related issues.*

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

| Code System | Data Protection | Decentralisation | Financial and Capital | Political and Institutional | Sovereignty | Security | SUM |
|---|---|---|---|---|---|---|---|
| Data Protection | | | | | | | 54 |
| Decentralisation | | | | | | | 27 |
| Financial and Capital | | | | | | | 35 |
| Political and Institutional | | | | | | | 55 |
| Sovereignty | | | | | | | 30 |
| Security | | | | | | | 19 |
| SUM | 54 | 27 | 35 | 55 | 30 | 19 | 220 |

***Figure 2.*** *Code relation matrix created from interview transcripts with all experts. Manually coded themes related to identifying problems and challenges were tagged, and their relationship with other themes are visualized based on how often they overlap, demonstrating how certain challenges are linked together. The most prominent relationships are political and institutional–financial and capital as well as political and institutional–data protection. Data protection issues also demonstrate some overlap with other problems and challenges more generally.*

that working with stakeholders of different backgrounds helped everyone better understand how a commons should be implemented and could be beneficial for reaching data protection goals. Our interview findings are addressed thematically below by each research question.

### *4.1. Identifying data protection challenges*

From the interviews, the experts identified data and data protection challenges related to the commons based on their own role-specific experiences. These challenges include the difficulty establishing the scope of the commons regarding data and data protection, assessing how the commons can be beneficial to those who participate, and determining the value of data included as well as the data protection benefits.

First, in identifying the data protection challenges within commons projects, interviewees mentioned that the main aims of the commons were often provided by the project coordinators. The experts themselves only had partial input on the scope of the commons and how the commons was to be defined. An interviewee in a technical role said that following their core commons aim: *"The most important challenge there was to make it decentralised"* (C1E1). Another interviewee elaborated that: *"Essentially what [the coordinators] wanted was, they realised that this [issue] poses a threat to [users'] privacy and they wanted us to build a system from the same dataset"* (C2E2). However, it was clear to some experts that data and data protection challenges would only more clearly emerge once the foundation of the commons was established alongside other stakeholders due to the nature of commons building. One interviewee said: *"The project was, we have these technologies, we do not know how these are going to be because we have not built it yet"* (C2E1). Another interviewee said: *"[One of the challenges is] striking a balance between openness and protection … and then just institutionalising that with advanced ICT"* (C2E3). As a result, what the precise scope of the commons is needs to be flexible to accommodate changes during the development process and incorporating participant input. This includes being open to changes when it comes to how data are collected and managed within the commons itself.

When discussing the benefits of a commons to data subjects for protecting their personal data, there is a role of responsibility from experts to communicate the options for protecting personal data: *"We played a role of coordination, and interaction with data subjects and data protection officers"* (C3E2). Another interviewee said that participants in a commons should understand that they have a real ability to have autonomy and sovereignty over their personal data, where the commons can support their preferences by operationalizing this control: *"in order for the data commons to work, you need to be able to give citizens some kind of control over their data, and give them, some kind of like, choice of what the data was going to be used for or not used for"* (C1E1). Beyond the challenges laid out by project coordinators, interviewees also mentioned that there were data protection challenges that go beyond the practical creation of the commons and included theoretical, philosophical, and psychological aspects of people's relationship with privacy. One interviewee summarized this eloquently: *"So essentially three challenges: Money and difficulty in the social side, distributing the technology, and the philosophical who owns divulged data in a*

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

*community side"* (C2E1). Wider socioeconomic issues surrounding data and Internet access also need to be addressed when considering and implementing a commons framework: *"The first thing I became aware of is the inequality in our access to the internet"* (C4E1). One interviewee suggested that rather than considering the commons framework as something put on top of a community, think about a data commons as intrinsically part of community collaboration: *"Digital space is infinite, we can have an infinitely large number of people in it, but we are still biological beings, we are still constrained by our biology and our grey matter up here. We can still only really build closed connections to this relatively small number of people. The question is not, in my opinion, how can we make commons all over the place, but more how can we bring together this biological and digital realities to optimise what is happening"* (C6E1). While the commons itself may be valuable in terms of increasing accessibility to data and knowledge, it must be managed in a way that is accessible and easily understood for data subjects in order for the commons to be successful.

In the period in which the interviews were conducted, the COVID-19 pandemic was taking place, and so in consideration of the data protection and wider data-related challenges, analogies related to the pandemic were used. One interviewee explained how tensions exist when it comes to building a commons and considering data protection through the public or private sector, challenging existing norms when it comes to the use of our personal data: *"We need to understand that we are giving all this information to the private sector to run our lives or to help us run our lives. This used to be delegated to the public sector so let us think about it or at least discuss about it, and see which model we really want because when something happens, like coronavirus nowadays, no one is looking for answers in the private sector but looking in the public sector. So I think a lot of reflection needs to be done in this and a lot of dialogue with the citizens and a lot of speech needs to be there"* (C2E2). In considering political and institutional barriers, one interviewee shared how a commons framework could be useful for opening up data and resources in a meaningful way: *"You have austerity destroying the public health infrastructure for a number of years, we have no valves, no ventilators, no masks, no protective equipment, and you see a mass of peer production groups that seek to solve these issues right? It is dialectic between the mainstream systems and increasing fault lines and then people self organising to find solutions beyond those bottlenecks basically"* (C3E1).

### *4.2. Overcoming data protection challenges*

According to the experts, establishing relationships with data subjects and developing trust in both the commons framework and those who created the framework was important for the commons' success, particularly regarding personal data and data protection. While many of the commoners were engaged with their specific projects, transparency and clarity in the process of contributing to the commons can foster an environment for engagement to achieve a better commons outcome for individuals and groups.

One aspect is creating trust and establishing positive relationships between those who have an understanding of the data commons and data protection with those who do not: *"The main problem was trying to be careful in understanding each other in achieving the goals but it was a cultural problem when you interact with different people from different backgrounds, and that's a problem you have working with different people"* (C3E2). Another aspect is bringing the community together within the commons. One interviewee said: *"Two things were really striking, the first one is this binary process where either the user trusts you or does not trust you. But once they trust you, they give you everything. This is the direct consequence of, you know when you accept the terms and conditions of the services, that's the same way"* (C2E1). Another interviewee further explained: *"Other than the legal constraints [surrounding data protection and privacy, we did not have any concerns that were raised]. This is one of the things that is really interesting and I think it is based on the trust. You have this social solidarity and there is this implicitly trust. If you break that trust, you are done"* (C6E1). This suggests that all stakeholders within the commons should feel that they are being respected and treated as experts bringing

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

in their own experiences, whether that may be knowledge, perspective, or personal anecdotes regarding their data.

Regardless of the use case of the commons, it is important to understand community concerns, applied both to data protection and other issues. For data protection, this includes recognizing the limitations of existing regulations and legal frameworks, such as the GDPR. One interviewee said: *"[Although, legally, you can ensure the process of deletion is followed,] you cannot tell people to forget something and they will forget. It was also something we realise with the GDPR law and our legal experts also discussed that"* (C2E1). These legal challenges regarding data protection also need to be considered throughout the commons development process, as Interviewee C3E2 explained the role of their team was to *"deal with legal issues related to the goals of the project, fostering the making of the digital commons including personal data."* In order to overcome these challenges, input is needed from the community to assess the benefits and risks to the use of their personal data. However, Interviewee C5E1 explained that although the community was willing to engage, they felt unable to do so, either because they did not know how or because they had been approached in a manner which did not appeal to them. The type of involvement related to the sharing of citizens' data-related worries and the data protection issues they were currently facing to enable their data protection rights to be enforced. Another interviewee explained that the commons framework is useful for unpacking the sociopolitical challenges that impact the community, rather than specifically seeking a technological solution: *"We have just used the term [data commons] to introduce [stakeholders] to this kind of thinking to immediately hear them out and how they feel about the data society, about the smart city discourse etcetera and see within their context, in mobility projects in certain neighbourhoods, or energy transition, how they feel they want to deal differently with these technologies and how with urbanity or neighbourhood initiatives"* (C2E3). As a result, when overcoming data protection challenges within a commons, it is important to acknowledge the limitations of the law, technologies, and data. The commons should support different methods for allowing data subjects to choose their own personal data protection preferences.

### *4.3. Improving the commons*

When discussing the usefulness and effectiveness of the commons, some interviewees expressed doubts. One said: *"I'm not entirely sure that [the project coordinators] actually achieved [their goals] in a reasonable sense because at some point there were too many challenges to resolve that and we took some short cuts in order to reasonably put something forward for the demo so there were lots of privacy issues that had to be solved later"* (C1E1), emphasizing the importance of timely development. Even in a commons, other stakeholders may be prioritized over data subjects, particularly when external financing and funding is involved: *"I often see the potential in people and areas in the project and then I have a hard line of what can and cannot be done and what the money was allocated for. So within our remit as an organization moving forward, it will be a huge conversation with the much higher ups than me about how do we deliver on our goals as set out in our original funding in a meaningful way that means that we are truly kind of, and I hope we have those conversations with people that are left out the most and working their way down to people who have access to things easily"* (C4E1). As a result, when establishing a successful commons, the scope of a specific commons is key in order to ensure that it is sustainable and balances the trade-offs between transparency and formalization with more fluid and iterative ways of working: *"One of the other risks that came about was this transfer from a small project to a bigger project. These like growing pains are always difficult and the new definition of roles, the formalisation of rules, is really really interesting and also when it starts to make money. When there starts to be something to have, something to gain, something that people want then the interpersonal relationships really change and that can be a real risk in particular with group cohesion"* (C6E1). As part of the development process, if there is no community consideration, policy can be negatively impacted. From an interviewee, over 60% from a group of 50,000 people surveyed had never been consulted before: *"It is very concerning at a policy level where we are trying to make consulting decisions based on what the community want or what the*

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

*stakeholders want or what the users want when the people we are hearing from are entirely unrepresentative of the local community"* (C5E1).

In order to improve the commons for data protection and overcome some of the identified challenges, all interviewees suggested that collaboration across stakeholders and disciplines could overcome excluding data subjects and doubts about the effectiveness of the commons. Working with stakeholders of different philosophical, technical, and social backgrounds helps everyone better understand how a commons should be implemented and could be beneficial for reaching data protection goals: *"I think the literacy gap will be always there. You cannot rely on the public money going to literacy and to train people in terms of technology or whatever so the delegation of trust and transparency are the key"* (C2E2). Another expert stressed the importance of inclusion: *"Low income and systemic inequality has left a lot of people not being able to access the internet like the rest of the world"* (C4E1). These considerations are also important when considering how data protection practices should be applied, on what mediums, and through what methods, particularly when addressing the reality on the ground one step at a time: *"The question for me is not how do we reach the end goal, the question is at the end of today, how are we one step closer to getting to the end goal at some point?"* (C6E1). Additionally, taking from the experience of COVID-19 and working online, some experts also believed that leaning into technologies and adopting digital tools in online–offline hybrid environments can make the commons more beneficial for a larger, more diverse group of people. One interviewee suggested: *"So it is one of those things that it is possible I think to develop social behavior in digital means and then to develop commons in that way…. I think the reason why I have a different perspective on this is that almost all of my work is remote so I have super close relationships with people in the UK, working relationships with people I have met once live, but we meet online once a week and we chat through all the stuff we are working on. I mean we are all freelancers, this is our way of meeting at the water cooler"* (C5E1). Another interviewee emphasized the importance of building connections in physical environments as well as digital ones: *"How can we flagship different connections and [not just] the transition to digital, … but I am talking about organizations, and by organizations I sometimes mean people who live on the street and want to set up planters, maybe communities or neighbours are better"* (C4E1). As a result, improving the commons requires direct community involvement where the means for co-creation best reflect how they interact with the commons and their data, both through online and offline means where appropriate.

### *4.4. Building a commons for data protection*

When considering the creation of a data protection-focused commons, there needs to be due consideration of the multidisciplinary nature of data protection and privacy as well as recognizing how the community's goals and how collective responsibility should be distributed.

One key point reiterated by many experts was the transition between theory and practice: *"Data commons is an idea that is hard to realise and our work was trying to make tangible example where this works and at least this had been tried and we'll see if it works or not"* (C2E1). The practicalities, given the strong relationship between issues related to data protection and wider sociopolitical environments, must not be tackled in a silo or within one discipline only: *"People need this commons perspective because they are thinking about open data and balancing the protection of data so we should use the value of collecting data and findings but at the same time seeing to the sovereignty of citizens. It is one thing to understand what does this look like but in practice, how can we operationalize this?"* (C2E3).

According to the experts, action and collective responsibility was also key. Another interviewee stressed the importance of action: *"[The commons] is a verb, it is commoning. It has the mindset of social solidarity and nonprofit oriented. It is democratic and nonhierarchical"* (C6E1). Several interviewees mentioned that the purpose of a data commons needs to be clear, as it is a choice. When building a data commons, more research needs to be done *"from legal, technical, social, political, economic areas of work"* and must include *"the vision of communities and people about what is at stake, what is this about,*

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

*how it works, [and] how [data] has been managed"* (C3E2). Importantly, individuals and communities need to be encouraged and empowered to co-create: *"A lot of people do commoning but they do not know they are commoning. They do not have an identity that permits them to have, to exert directly power"* (C3E1).

Finally, it is important to consider what the goals of the community are and addressing those directly and collectively: *"I think that things will evolve because you know in free software, things evolved during decades from the beginning until now. Now things are very different, the vision of communities and people about what is at stake, what is this about, how it works, how it has been managed has changed a lot. I think something similar is happening with data now. We have the issue of tracking, contagion tracking apps, you know. This is creating a lot of debate about for example the data, why should the data be used for benefiting the community that is connected?"* (C3E2). Another interviewee explained: *"So it is related to those aspects, what do you give, what do you contribute to the commons and how will it be used? It is more on that issue that I have concerns. Who is deciding on how the commons is going to be used?"* (C1E1). Ultimately, as framed by one interviewee, the commons framework is seen as an alternative way of considering how a resource could be managed through transparently communicating risk and offering adequate protection within different hierarchical norms and rules as determined by the community itself: *"Whereas if you collectivise that risk, into an organization like a union, then that body is about the same size as this other organization and it is the same with commons. You are collectivizing risk and you are also collectivizing benefits because everything is distributed among the group in an equal way"* (C6E1). In this way, a data protection-focused data commons can support the community in making the most out of their data and personal data, without disregarding the importance of protecting that data in the first place.

### *4.5. Summary of interviews*

Our interviews indicated that data protection within existing commons frameworks was predominantly considered only in terms of control and sovereignty over data subjects' personal data. Although the decision to use a commons was to provide certain levels of control and transparency of how data were collected, used, and processed, there were limited applications to wider data protection principles such as those relating to informing data subjects about their rights and the ability to exercise those rights against data controllers. According to the experts, establishing strong community relationships to develop trust in both the commons framework itself and those who created the framework was important for the commons' success. While many of the commoners were engaged with their specific projects, transparency and clarity in the process of contributing to the commons can foster an environment for engagement in order to achieve a better commons outcome for both individuals and groups. Openly acknowledging the legal and technological limitations within data protection can also result in a more supportive and collaborative environment for creating data protection solutions. Interviewees also expressed their doubts over the use of the commons framework for their specific projects, as certain assumptions were made about its ability to be put into practice. However, all interviewees suggested that collaboration through online and offline means across disciplines could overcome this challenge. The commons can support data protection and allow data subjects to extract value from the knowledge generated from their data without sacrificing privacy where the limitations of legal protections or technological innovation are communicated. Working with stakeholders of different philosophical, technical, and social backgrounds helps everyone better understand how a commons should be implemented and could be beneficial for reaching data protection goals.

## 5. Discussion

In considering creating a data protection-focused data commons, the experts identified important considerations throughout the development process. While their experiences highlight the importance of including data subjects early on in creating a commons, there remains open questions about how to implement and develop a data protection-focused data commons. Using the findings from the interviews,

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

we thematically elaborate on how the challenges identified by the interview experts can be overcome. We then adapt these themes and solutions to the IAD framework, transcribing the analytical framework into a practical policy checklist to support commons policy-makers in developing data protection-focused data commons that supports genuine and adequate engagement.

### *5.1. Adopting a data protection-focused data commons*

In order to overcome the challenges of creating a commons and adopting the framework as a form of data stewardship and management, based on the expert interviews, we identify important themes of consideration for adoption of a data protection-focused data commons: multidisciplinary solution building, accessibility, as well as community and social solidarity.

#### *5.1.1. Multidisciplinary solution building*

First, a commons is useful due to multidisciplinary solution building. As noted from the interviews, as with many data stewardship methods, commons frameworks are not one-size-fit-all solutions for data protection-related issues. Rather than focusing on adapting specific legal doctrines or building new independent technological infrastructures, the data protection-focused data commons incorporates the principles and spirit of data protection law through data infrastructures already used by the community for protecting individuals' personal data. The commons also does not require registration or incorporation as a separate legal entity and can be adopted within existing organizations and infrastructures. This means that the strengths and weaknesses of current laws, technologies, and norms can be identified and improved using solutions from different disciplines, thus increasing the flexibility and adaptability of the commons for supporting data subject data protection preferences. Data subjects' personal experience in both data protection and the norms of their community also plays an important role in understanding what the data protection challenges are and can apply that knowledge directly into the commons.

#### *5.1.2. Accessibility*

Additionally, accessibility considerations are also an important part of overcoming commons challenges. As mentioned by experts, people often do not know that they are commoning or do not think they are able to participate in decision-making. This can be a result of a lack of transparency of the aims and objectives of a commons, not adequately seeking participant input, or not creating infrastructures that can be easily accessed by commoners. In order to ensure that a commons can benefit the data subjects who are contributing to it, the commons must be created in an environment with as little friction as possible. This means that the commons itself should be technological accessible to the group, where existing community norms and rules should be preserved on that platform. The data protection-focused commons itself should extend the knowledge commons, where information is not only shared, but also explained and applied to ensure that data subjects are able to make the most out of commons resources. Contributions to the commons should be encouraged while acknowledging the difficulties individuals may be facing with regard to their ability to connect to the platform and the personal sensitivities of discussing personal data issues. As a result, part of adopting a data protection-focused data commons involves addressing accessibility challenges to understanding and protecting data within a commons environment.

#### *5.1.3. Community and social solidarity*

Finally, an important aspect of overcoming commons challenges is by continuing to develop a sense of community for the specific aims of the commons. Taking the examples of open-source software and physical neighborhoods as identified by experts, the usefulness of the commons comes from a sense of identity and commune. As there are many forms in which a commons could take place, individuals and groups of data subjects can establish their own collective sense of belonging and not be dictated by their data. This is particularly important for when unexpected uses of data occur. By being able to actively participate in a commons, data subjects not only benefit from a data protection perspective, but can also

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

see how their personal data represent the community and their information are not just reduced to numbers. Their group involvement in the design and creation of the commons can help collectivize the risks and benefits of data sharing, thereby reducing the responsibilization of the data protection process. Building social solidarity through a commons, as opposed to having to solve data protection problems yourself or relying on a third party to act for you, allows data subjects to regain their agency over their personal data as well as themselves as individuals.

By considering these themes right from the beginning of creating a commons, the commons can support the transparent communication of personal data risks within community determined norms and rules while offering the protection of personal data both during the data management process as well as supporting the remedy of potential breaches after the personal data have been curated.

### *5.2. Data commons checklist*

Taking our findings and adapting existing theories on the knowledge commons framework to develop a data protection-focused data commons, we apply the interview findings, themes, and questions considered by the experts from their interviews identified from Section 4 to the IAD framework as discussed in Section 2. This is included as part of the Supplementary Material of this article. We also address the commons policy implications noted by Sanfilippo et al. (2018) and elaborate on how policy-makers can promote appropriate information flows while protecting personal data.

Given the IAD framework's focus on the analysis of a commons rather than the practical development of a commons resource, in order to better support practitioners in meeting their policy agendas when developing the commons, we establish a checklist for commons policy-makers based on the themes outlined by the experts. The data commons policy checklist covers the same content of the modified IAD framework but focuses on the scope and impact the commons intends to have to clarify how the commons' aims could fit into wider policy aims. Establishing policy documentation and creating an action items checklist for policy-makers encourages data protection to be more holistically considered for each use case to allow for co-creation, engagement, and participation for all stakeholders within a commons. The key aspects of implementing a commons can be identified and more easily put into practice by policy-makers and applied to specific use cases for local communities. As the process of creating a commons requires having data subject participation from the beginning, this checklist places their considerations throughout the commons development process by directly asking policy-makers questions that can only be answered by data subject and so require that they be included in the process. Taking the analysis questions developed by McGinnis (outlined in Section 2) and incorporating them into this checklist, this set of analysis questions is more suitable for creating a policy checklist, as it provides a more holistic view of the commons development process beyond the IAD framework while also including direct data subject engagement. Given that the IAD framework is not a legislative proposal, the framework should be considered a means to initiate and use the commons within existing legal and sociotechnical infrastructures to encourage engagement and participation. The contents of the checklist are split into four parts, mirroring the interviews analysis in Section 4: identifying the commons use case and the data subjects, scoping and information gathering for developing the commons, building the commons, and sustaining the commons. Each part reflects the themes drawn out from our interviews, particularly on how a commons can be improved for data protection. The checklist also incorporates advice from experts on how to meaningfully involve data subjects at each stage of development, ensuring that the aim of the commons supports data subjects in creating data protection solutions. The checklist is as follows.

#### *5.2.1. Identifying the commons use case and the data subjects*

The first step for policy-makers when creating a data protection-focused data commons is to identify the use case to which a commons can be applied and who are the data subjects the commons is aimed to benefit. When examining a data commons use case, a data protection-focused data commons could serve as a new public consultation mechanism for policy-makers and help identify data protection best practices to incorporate into policy. Directly incorporating data commons policies into consultation work allows

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

data use, sharing, and methods for protection to be transparent, ensuring that their perspectives are considered in the process:

- What is the data protection issue for the data subjects for the use case?
- Is the issue one that relates more to finding a policy solution or does it require a wider scope in identifying an underlying problem? What could have caused this issue and was there any event that may have exemplified it?
- What resources are already available to support data subjects and how should this information be presented to them within a commons to make it more accessible?
- Which data subjects would be invited to participate and engage in the public consultation process?
- How should data subjects be included?
- What value does better data protection for this use case bring to the data subjects involved and also to the data itself?

#### *5.2.2. Scoping and information gathering for developing the commons*

The next step to consider involves stakeholders, including data subjects early on in the process. For data subjects, when creating and using a data commons, writing new community policies as well as using existing data protection policies, such as regulations and institutional policies or codes, can support them in co-creating data protection responsibilities for and alongside other stakeholders. Guidance should also be provided for data subjects should they wish to co-create policies within the data commons:

- Who are the other stakeholders with more power over personal data compared to the data subjects?
- Who are the other stakeholders that can support data subjects and provide more information for them?
- What relationships do these stakeholders have between each other? Are there specific stakeholders that are dominating what happens within the commons use case identified?
- How would information, advice, and participation from different stakeholders, including data subjects, be included during the commons development process?
- What are the wider data protection and privacy issues (social, technological, and philosophical) that relate to this use case?
- What solutions, if there are any at all, have data subjects tried in an attempt to solve the data protection issue identified?
- Are there similar commons or data stewardship examples that can help support the creation of this new commons?
- What laws, technologies, or policies have been developed for this use case that could support better data protection practices? Are these enforced under the law, codes or conduct, or the community? Are these effective?

#### *5.2.3. Building the commons*

Once the stakeholders have been identified and a preliminary blueprint has been drafted, the next step involves creating and building the commons by addressing some of the issues previously identified in more depth. This stage involves more involvement to ensure that the commons development process can be iterative and best reflect data subject preferences. Some of these questions include:

- As part of the commons process, how will you find out what data subjects' data-related worries are and how to support their data protection rights?
- How can the wider data protection and privacy issues identified previously be addressed either as part of the commons itself or from the commons development process?
- Within the commons, what mechanisms can help develop trust between stakeholders, particularly for data subjects?

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

- What assumptions related to the use case need to be addressed and corrected? What baseline information should data subjects have to best help them co-create the most suitable data protection solution?

#### *5.2.4. Sustaining the commons*

Finally, policy-makers should consider how the commons can operate in the long term with other stakeholders as well. For example, it may be useful to include data controllers, so that they can better understand what data protection requirements are preferred by data subjects. For data protection authorities, policies established around creating a data commons for specific use cases help ease their burden of enforcement through preventative data protection measures, ex ante, before data are collected as opposed to remedying data breaches, ex post. Additionally, establishing a data protection-focused data commons framework in policy encourages policy-makers to reconsider current balances of power between data subjects, data controllers, and other data protection stakeholders, taking into consideration the data ecosystem in the long term for socioeconomic benefit by increasing the value of data:

- How can the commons be sustainable in the long run? What can be done at the development and implementation stage to ensure that the data protection issue can be better managed and solved?
- Are there particular platforms or infrastructures that can help host the commons and ensure that it is as accessible as possible?
- How can collective responsibility be demonstrated within the commons and how can its reach be maximized?
- What other stakeholders can be brought in to help support and sustain the commons?
- How will you know when the data protection goal of the commons is achieved? How can this be measured?

## 6. Future Work

Based on our interviews, we have found the importance of including data subjects in the commons development process, where multidisciplinary considerations for data protection need to be taken to support data subjects' data protection goals. Building the commons includes recognizing the theoretical, philosophical, and psychological aspects of people's relationship with privacy and allowing those in the commons to be able to express their views. These perspectives should be included as part of the commons' goal based on the use case. As the commons has been identified as a means of community collaboration, opportunities for feedback and iteration, such as chats, forums, and public consultation processes, need to be included in the commons process. Importantly, data protection within the commons itself must also be transparent and reflect the needs of data subjects.

### *6.1. Data commons in practice*

With a framework for developing a data protection-focused data commons, the next step involves testing the practicality of applying data protection to a commons against its usefulness for data subjects in projecting their data protection preferences. Given the current shift toward online teaching and remote learning, we will be creating a data commons tool to be tested for this particular use case. Based on the experts' perspective on how to create trust and community within a commons, in creating the application, interactive forums and means of communicating both within the commons community as well as with external experts have been included. From the interviews, action and collective responsibility were identified as a core part of the commons. This is represented through encouraging those in the commons to share their data protection experiences with each other through the tool and including their visions of what they hope the application can help them with in the long term. Specific elements of the data commons to be tested include building opt-in mechanisms within existing platform to test whether these tools encourage data subjects to make better data protection choices,

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

assessing whether having access to other data protection materials, sources, and information within a commons helps data subjects better understand the data protection options, and if prompting data subjects to exercise their data protection rights may encourage them to learn about how their personal data are being used by data controllers.

### *6.2. Deploying a data protection-focused data commons*

In order to deploy a data protection-focused data commons, based on our findings, nontechnical language should be used, removing the barrier for engagement and co-creation. While some interviewees acknowledged that the data and digital literacy gap will remain, it is important that inclusive data infrastructures are considered when protecting our personal data. Within the commons, an expert or a person with more knowledge on how data are used should be included in the commons, allowing data subjects to direct their concerns to them.

Other technological aspects of personal data that need to be considered include the significance of metadata. Particularly with regard to data protection, metadata plays an important role as to how much information data subjects would like to share with other stakeholders. By considering a data protection-focused data commons as a tool for facilitating data portability and interoperability, data subjects may be able to better understand how their personal data can be used in different contexts, particularly where there are currently limited guidance on how data portability could be enforced in the context of data protection (Wong and Henderson, 2019) and how such expectations are constantly changing (Li, 2021). The consideration of data portability and interoperability within a data commons also go beyond existing data stewardship perspectives that focus on data ownership by including data experimentation and interaction (Mills, 2019). As a result, data infrastructures need to be assessed more broadly, continuing the conversation of how a data protection-focused data commons could contribute to better data stewardship and data governance. Based on the experts' advice, the commons should be as integrated into the existing data infrastructure as possible, minimizing the friction between using the commons and managing the data protection issue itself. For example, if a commons tool was to be created for online learning, it could sit on top of the same platform that the online learning is taking place on.

To deploy a data commons in the long term, as mentioned by the experts, consideration needs to be made with regard to the platform used to host the commons and how the commons is to be sustained financially. Given the difference in stakeholder interests, how and by whom the commons is maintained can impact the trust between users as data subjects and others that participate in the development of the commons. Within the commons itself, experts suggest that these decisions need to be made collectively by data subjects with support from stakeholders they believe to be supporting them, which can change over time and will differ from use case to use case. The commons should therefore have dedicated periods of review. For example, in the case of online learning, this could be once every academic year, to ensure that the goal of the commons reflects the features and functionality of the commons itself. By considering data protection itself as a commons process, existing barriers to access and power imbalances can begin to break down, as different stakeholders are more transparent about data collection, use, and processing.

Finally, in tangent with deploying a data protection-focused data commons, wider conversations need to be considered regarding how personal data are being treated more broadly. This includes discussing the impact data regulations and policies on data subjects. For example, within data protection regulation, access to the fundamental right to data protection through the exercise of data rights can be further strengthened (Ausloos et al., 2020). Laws such as the European Data Governance Act which aims to increase trust in data intermediaries and strengthen data-sharing mechanisms across the EU could also support better data protection practices for the benefit of data subjects outside of data protection. Research and guidance from organizations and advisory bodies such as the Centre for Data Ethics and Innovation in the United Kingdom can also play an important role in connecting different stakeholders and addressing data issues to specific domains. By furthering discussions around data governance in support of data protection, a data protection-focused data commons can not only address data subject issues, but also take into account the bigger picture in relation to how personal data can be protected for the common good.

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

## 7. Conclusion

In this paper, we set out how a data protection-focused data commons can support more accountable data protection practices, management, and sharing for the benefit of data subjects, data controllers, and policy-makers to overcome the limitations of laws and technologies in protecting personal data. Although existing data stewardship frameworks aim to increase the value of data sharing through mitigating data-related harms and the transfer of rights, data subjects themselves may not be directly supported, particularly where current models do not help remedy potential data breaches using data rights. Adopting existing commons frameworks, personal data can be better protected through co-creation and collaboration with data subjects, placing their data protection preferences at the center of the decision-making process. From our interviews with commons experts, we identified the data protection challenges for creating a commons, how to overcome them, how to improve the commons more broadly, and the important requirements for building a data protection-focused data commons. Based on those themes, we adapted the IAD framework for data protection to support the deployment of a data protection-focused data commons in practice and created a checklist for policy-makers to allow them to apply the commons to specific use cases, outlining key questions that should be answered at each step of the commons development process. We suggest working with stakeholders of different backgrounds and perspectives at an early stage of the commons development to support its implementation, proposing further considerations necessary for deploying a data protection-focused data commons in practice. By applying the data protection-focused data commons and developing policies on its implementation, data protection can be improved as a common good, mitigating the power imbalances between data subjects and other stakeholders when it comes to personal data.

**Acknowledgments.** We would like to thank our reviewers for their time and feedback. We would also like to thank our interviewees for their participation. A shorter version of this article was presented at the Data for Policy 2020 Conference: https://doi.org/10.5281/zenodo.3965670.

**Funding Statement.** This research is part of Janis Wong's doctoral research, which is funded by the University of St Andrews St Leonard's College, School of Computer Science, and School of Management.

**Competing Interests.** The authors declare no competing interests exist.

**Author Contributions.** Conceptualization: J.W., T.H., and K.B.; Methodology: J.W., T.H., and K.B.; Investigation: J.W.; Data curation: J.W.; Data visualization: J.W.; Supervision: T.H. and K.B.; Writing—original draft: J.W.; Writing—review and editing: J.W., T.H., and K.B. All authors approved the final submitted draft.

**Data Availability Statement.** As data from the interviews are presented anonymously, no interview transcripts have been made publicly available. A document with the semi-structured interview questions has been provided in the Supplementary Material.

**Ethical Standards.** The research meets all ethical guidelines, including adherence to the legal requirements of the study country. This study has been granted ethical approval (approval code CS14765) by the School of Computer Science Ethics Committee on behalf of the University Teaching and Research Ethics Committee at the University of St Andrews.

**Supplementary Materials.** To view supplementary material for this article, please visit http://dx.doi.org/10.1017/dap.2021.40.

## References

**Ada Lovelace Institute** (2021a) Data cooperatives. Available at https://www.adalovelaceinstitute.org/feature/data-cooperatives/ (accessed 26 August 2021).

**Ada Lovelace Institute** (2021b) Data trusts. Available at https://www.adalovelaceinstitute.org/feature/data-trusts/ (accessed 26 August 2021).

**Ada Lovelace Institute** (2021c) Exploring legal mechanisms for data stewardship. Available at https://www.adalovelaceinstitute.org/wp-content/uploads/2021/03/Legal-mechanisms-for-data-stewardship_report_Ada_AI-Council-2.pdf (accessed August 26 2021).

**Ada Lovelace Institute** (2021d) Exploring principles for data stewardship—a case study analysis. Available at https://docs.google.com/spreadsheets/d/1hAN8xMJuxobjARAWprZjtcZgq1lwOiFT7hf2UsiRBYU/edit#gid=432908716 (accessed 26 August 2021).

**Andersen IE and Jæger B** (1999) Scenario workshops and consensus conferences: towards more democratic decision-making. *Science and Public Policy 26*(5), 331–340. https://doi.org/10.3152/147154399781782301. eprint available at https://academic.oup.com/spp/article-pdf/26/5/331/4685076/26-5-331.pdf

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

**Ausloos J**, **Mahieu R and Veale M** (2020) Getting data subject rights right a submission to the European data protection board from international data rights academics, to inform regulatory guidance. *Journal of Intellectual Property, Information Technology and E-Commerce Law 10*(3), 283–309. Available at http://nbn-resolving.de/urn:nbn:de:0009-29-50315

**Australia Research Data Commons** (2020) Australia Research Data Commons. Available at https://ardc.edu.au (accessed 26 August 2021).

**Benkler Y**, **Shaw A and Hill MB** (2015) Peer production: a modality of collective intelligence. In Malone TW and Bernstein MS (eds), *Handbook of Collective Intelligence*. Cambridge, MA: MIT Press, pp. 175–204.

**Bennett Institute for Public Policy and the Open Data Institute** (2020) The value of data summary report 2020. Available at https://www.bennettinstitute.cam.ac.uk/media/uploads/files/Value_of_data_summary_report_26_Feb.pdf (accessed 26 August 2021).

**Bergold J and Thomas S** (2012) Participatory research methods: a methodological approach in motion. *Forum Qualitative Sozialforschung 13*(1). Available at http://nbn-resolving.de/urn:nbn:de:0114-fqs1201302

**Bolychevsky I** (2021) How solid is Tim's plan to redecentralize the web? Available at https://medium.com/zero-equals-false/how-solid-is-tims-plan-to-redecentralize-the-web-b163ba78e835 (accessed 4 October 2018).

**Bria F** (2021) The EU must be bold and defend its digital sovereignty. Available at https://www.ft.com/content/84dbe3a0-3a40-43bd-850d-ba8e3cab34cd (accessed 26 August 2021).

**Cadwalladr C and Graham-Harrison E** (2018) Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. Available at https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election (accessed 26 August 2021).

**California State Legislature** (2018) The California Consumer Privacy Act of 2018. *California Legislative Information 375*, 1–24. Available at https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375

**Cecco L** (2020) Google affiliate Sidewalk Labs abruptly abandons Toronto smart city project. Available at https://www.theguardian.com/technology/2020/may/07/google-sidewalk-labs-toronto-smart-city-abandoned (accessed 26 August 2021).

**Cellan-Jones R** (2020) Coronavirus: England's test and trace programme "breaks GDPR data law." Available at https://www.bbc.com/news/technology-53466471 (accessed 26 August 2021).

**Cisco Secure 2020 Consumer Privacy Survey** (2020) Protecting data privacy to maintain digital trust: the importance of protecting data privacy during the pandemic and beyond. Available at https://www.cisco.com/c/dam/en_us/about/doing_business/trust-center/docs/cybersecurity-series-2020-cps.pdf (accessed 26 August 2021).

**Coulter M** (2021) Alphabet's Sidewalk Labs has abandoned another US smart city project after reported fights about transparency. Available at https://www.businessinsider.com/second-sidewalk-labs-smart-city-project-shutters-portland-oregon-2021-2?r=US&IR=T (accessed 26 August 2021).

**Coyle D** (2020) Common governance of data: appropriate models for collective and individual rights. Available at https://www.adalovelaceinstitute.org/blog/common-governance-of-data/ (accessed 26 August 2021).

**Crabtree A**, **Lodge T**, **Colley J**, **Greenhalgh C**, **Mortier R and Haddadi H** (2016) Enabling the new economic actor: data protection, the digital economy, and the Databox. *Personal and Ubiquitous Computing*. https://doi.org/10.1007/s00779-016-0939-3

**Custers B**, **Sears AM**, **Dechesne F**, **Georgieva I**, **Tani T and van der Hof S** (2019) Conclusions. In *EU Personal Data Protection in Policy and Practice*, pp. 195–233. https://doi.org/10.1007/978-94-6265-282-8_10

**Cuthbertson A** (2021) WhatsApp privacy controversy causes "largest digital migration in human history," Telegram boss says as he welcomes world leaders. Available at https://www.independent.co.uk/life-style/gadgets-and-tech/whatsapp-privacy-telegram-world-leaders-b1787218.html (accessed 26 August 2021).

**Data Economy Lab** (2021a) Data cooperative. Available at https://tool.thedataeconomylab.com/data-models/1 (accessed 11 November 2021).

**Data Economy Lab** (2021b) Data stewardship models. Available at https://tool.thedataeconomylab.com/our-data-models (accessed 11 November 2021).

**Data Economy Lab** (2021c). Data trust. Available at https://tool.thedataeconomylab.com/data-models/10 (accessed 11 November 2021).

**Data Trusts Initiative** (2021) Seeking data trusts pioneers! Funding from the data trusts initiative will support pilot projects to set up real-world data trusts. Available at https://datatrusts.uk/blogs/seeking-data-trusts-pioneers-funding-from-the-data-trusts-initiative-will-support-pilot-projects-to-set-up-real-world-data-trusts (accessed 26 August 2021).

**De Marchi B** (2003) Public participation and risk governance. *Science and Public Policy 30*(3), 171–176. https://doi.org/10.3152/147154303781780434. eprint available at https://academic.oup.com/spp/article-pdf/30/3/171/4602581/30-3-171.pdf

**Decidim** (2021) Decidim. Available at https://decidim.org/ (accessed 26 August 2021).

**DECODE** (2020) DECODE application. Available at https://github.com/DECODEproject/decode-app (accessed 26 August 2021).

**Delacroix S and Lawrence ND** (2019) Bottom-up data trusts: disturbing the "one size fits all" approach to data governance. *International Data Privacy Law 9*. https://doi.org/10.1093/idpl/ipz014

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

**Dellenbaugh-Losse M**, **Zimmermann N-E and de Vries N** (2020) *The Urban Commons Cookbook: Strategies and Insights for Creating and Maintaining Urban Commons*. La Vergne, TN: IngramSpark.

**Dourish P and Anderson K** (2006) Collective information practice: exploring privacy and security as social and cultural phenomena. *Human–Computer Interaction 21*(3), 319–342. https://doi.org/10.1207/s15327051hci2103_2

**Driver's Seat** (2020) Driver's Seat. Available at https://driversseat.co/

**European Commission** (2018) Reclaiming the smart city: personal data, trust and the new commons. *DECODE.* Available at https://media.nesta.org.uk/documents/DECODE-2018_report-smart-cities.pdf

**European Commission** (2019) European Open Science Cloud (EOSC) strategic implementation plan. *European Commission, 01*. Available at https://op.europa.eu/en/publication-detail/-/publication/78ae5276-ae8e-11e9-9d01-01aa75ed71a1/language-en

**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*, 1–88. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC

**Fiesler C and Hallinan B** (2018) "we are the product": public reactions to online data sharing and privacy controversies in the media. In *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–13. https://doi.org/10.1145/3173574.3173627

**Fisher JB and Fortmann L** (2010) Governing the data commons: policy, practice, and the advancement of science. *Information & Management 47*(4), 237–245. https://doi.org/10.1016/j.im.2010.04.001

**Fung A** (2015) Putting the public back into governance: the challenges of citizen participation and its future. *Public Administration Review 75*(4), 513–522. https://doi.org/10.1111/puar.12361. eprint available at https://onlinelibrary.wiley.com/doi/pdf/10.1111/puar.12361

**gE.CO** (2021). gE.CO Living Lab about page. Available at https://generative-commons.eu/ (accessed 26 August 2021).

**Governance Lab** (2021a) Data collaboratives. Available at https://datacollaboratives.org/ (accessed 26 August 2021).

**Governance Lab** (2021b) Wanted: data stewards: (re-)defining the roles and responsibilities of data stewards for an age of data collaboration. Available at https://www.thegovlab.org/static/files/publications/wanted-data-stewards.pdf (accessed 26 August 2021).

**Grossman RL**, **Heath A**, **Murphy M**, **Patterson M and Wells W** (2016) A case for data commons: toward data science as a service. *Computing in Science Engineering 18*(5), 10–20. https://doi.org/10.1109/MCSE.2016.92

**Hardinges J** (2020) Data trusts in 2020. Available at https://theodi.org/article/data-trusts-in-2020/ (accessed 26 August 2021).

**Herrmann M**, **Hildebrandt M**, **Tielemans L and Diaz C** (2016) Privacy in location-based services: an interdisciplinary approach. *SCRIPTed 13*. https://doi.org/10.2966/scrip.130216.144

**Hess C** (2006) Research on the commons, common-pool resources, and common property. *Indiana University Digital Library of the Commons.* Available at http://dlc.dlib.indiana.edu/dlc/contentguidelines

**Hess C and Ostrom E** (2007) *Understanding Knowledge as a Commons: From Theory to Practice*. Cambridge, MA: MIT Press.

**Hill K and Mattu S** (2018) The house that spied on me. Available at https://gizmodo.com/the-house-that-spied-on-me-1822429852 (accessed 26 August 2021).

**Human-Centered Artificial Intelligence, Stanford University** (2021) Radical proposal: data cooperatives could give us more power over our data. Available at https://hai.stanford.edu/news/radical-proposal-data-cooperatives-could-give-us-more-power-over-our-data (accessed 11 November 2021).

**International Cooperative Alliance** (2018) Cooperative identity, values & principles. Available at https://www.ica.coop/en/cooperatives/cooperative-identity (accessed 1 November 2021).

**Involve** (2019) Designing decision making processes for data trusts: lessons from three pilots. Available at https://www.involve.org.uk/sites/default/files/field/attachemnt/General-decision-making-report-Apr-19.pdf (accessed 26 August 2021).

**Jensen MA**, **Ferretti V**, **Grossman RL and Staudt LM** (2017) The NCI genomic data commons as an engine for precision medicine. *Blood 130*(4), 453–459. https://doi.org/10.1182/blood-2017-03-735654

**Jumbo Privacy** (2019) Jumbo privacy. Available at https://www.jumboprivacy.com/ (accessed 26 August 2021).

**Lafontaine E**, **Sabir A and Das A** (2021) Understanding people's attitude and concerns towards adopting IoT devices. In *CHI EA '21: Extended Abstracts of the 2021 Chi Conference on Human Factors in Computing Systems.* https://doi.org/10.1145/3411763.3451633

**Laziuk E** (2021). iOS 14.5 opt-in rate—daily updates since launch. Available at https://www.flurry.com/blog/ios-14-5-opt-in-rate-att-restricted-app-tracking-transparency-worldwide-us-daily-latest-update/ (accessed 26 August 2021).

**Lee D**, **Levi M and Brown JS** (2021) Democratic societal collaboration in a whitewater world. In Bernholz L, Landemore H and Reich R (eds), *Digital Technology and Democratic Theory*, pp. 219–240. https://doi.org/10.7208/chicago/9780226748603.001.0001

**Lewis P** (2020). Peter Lewis's 2020s vision: stop glibly signing over your data and take control. Available at https://www.theguardian.com/commentisfree/2020/feb/22/peter-lewiss-2020s-vision-stop-glibly-signing-over-your-data-and-take-control (accessed 26 August 2021).

**Li W** (2021) Between incrementalism and revolution: how the GDPR right to data portability is revamped by the EU and the UK post Brexit. https://doi.org/10.31219/osf.io/8u2pr

**Mahieu R**, **Asghari H and van Eeten M** (2017) Collectively exercising the right of access: individual effort, societal effect. In *GigaNet (Global Internet Governance Academic Network) Annual Symposium 2017.* https://doi.org/10.2139/ssrn.3107292

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

**Mahieu R and Ausloos J** (2021) Harnessing the collective potential of GDPR access rights: towards an ecology of transparency. Available at https://policyreview.info/articles/news/harnessing-collective-potential-gdpr-access-rights-towards-ecology-transparency/1487 (accessed 26 August 2021).

**Mansour E**, **Sambra AV**, **Hawke S**, **Zereba M**, **Capadisli S**, **Ghanem A**, **Aboulnaga A**, **Berners-Lee T** (2016) A demonstration of the solid platform for social web applications. In *WWW '16 Companion: Proceedings of the 25th International Conference Companion on World Wide Web*, pp. 223–226. https://doi.org/10.1145/2872518.2890529

**McDonald S** (2021) Data governance's new clothes. Available at https://www.cigionline.org/articles/data-governances-new-clothes/ (accessed 1 November 2021).

**McGinnis MD** (2018) The IAD framework in action: understanding the source of the design principles in Elinor Ostrom's governing the commons. In Coleand D and McGinnis MD (eds), *Elinor Ostrom and the Bloomington School of Political Economy, Volume 3: A Framework for Policy Analysis*. Lanham: Lexington, pp. 87–108. Available at https://polisci.indiana.edu/documents/profiles/mcginnis1.pdf

**Mills S** (2019) Who owns the future? Data trusts, data commons, and the future of data ownership. https://doi.org/10.2139/ssrn.3437936

**National Cancer Institute** (2020) Genomic Data Commons. Available at https://gdc.cancer.gov/ (accessed 26 August 2021).

**New Economics Foundation** (2018) Co-operatives unleashed: doubling the size of the UK's co-operative sector. Available at https://neweconomics.org/uploads/files/co-ops-unleashed.pdf (accessed 11 November 2021).

**Nissenbaum H** (2004) Privacy as contextual integrity. *Washington Law Review 79*, 119–158. Available at https://heinonline.org/HOL/Page?handle=hein.journals/washlr79&id=129&collection=journals&index=

**Norris C**, **de Hert P**, **L'Hoiry X and Galetta A** (eds) (2017) *The Unaccountable State of Surveillance*. https://doi.org/10.1007/978-3-319-47573-8

**O'Hara K** (2019) Data trusts: ethics, architecture and governance for trustworthy data stewardship. Available at https://eprints.soton.ac.uk/428276/1/WSI_White_Paper_1.pdf (accessed 26 August 2021).

**Open Data Institute** (2019) Data trusts: lessons from three pilots. Available at https://docs.google.com/document/d/118RqyUAWP3WIyyCO4iLUT3oOobnYJGibEhspr2v87jg/edit (accessed 26 August 2021).

**Open Data Institute** (2021) What are data institutions and why are they important? Available at https://theodi.org/article/what-are-data-institutions-and-why-are-they-important/ (accessed 26 August 2021).

**Open Usage** (2021) Open Usage Commons. Available at https://openusage.org/ (accessed 26 August 2021).

**OpenGDPR** (2018) OpenGDPR. Available at https://github.com/opengdpr/opengdpr (accessed 26 August 2021).

**Ostrom E** (1990) *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge, UK: Cambridge University Press.

**Ostrom E** (2005). *Understanding Institutional Diversity*, 1st Edn. Princeton, NJ: Princeton University Press.

**Ostrom E** (2010) Polycentric systems for coping with collective action and global environmental change. *Global Environmental Change 20*, 550–557. https://doi.org/10.1016/j.gloenvcha.2010.07.004

**Ostrom E** (2012) *The Future of the Commons: Beyond Market Failure & Government Regulations*. London: Institute of Economic Affairs.

**Ostrom V**, **Tiebout CM and Warren R** (1961) The organization of government in metropolitan areas: a theoretical inquiry. *American Political Science Review 55*, 831–842.

**P2P Foundation Wiki** (2021). Data cooperatives. Available at https://wiki.p2pfoundation.net/Data_Cooperatives (accessed 26 August 2021).

**Peppin A** (2020). Doing good with data: what does good look like when it comes to data stewardship? Available at https://www.adalovelaceinstitute.org/doing-good-with-data-what-does-good-look-like-when-it-comes-to-data-stewardship/ (accessed 26 August 2021).

**Perrin A** (2020) Half of Americans have decided not to use a product or service because of privacy concerns. Available at https://www.pewresearch.org/fact-tank/2020/04/14/half-of-americans-have-decided-not-to-use-a-product-or-service-because-of-privacy-concerns/ (accessed 26 August 2021).

**Powell M** (2014) What's the difference between a foundation and a trust? Available at https://www.hawksford.com/knowledge-hub/2014/foundations-vs-trusts (accessed 26 August 2021).

**Rabley P and Keefe C** (2021) Establishing a data trust: it's really hard. Available at https://www.thisisplace.org/blog-1/introducing place/its-really-hard (accessed 26 August 2021).

**Sanfilippo M**, **Frischmann B and Standburg K** (2018) Privacy as commons: case evaluation through the governing knowledge commons framework. *Journal of Information Policy 8*, 116–166. Available at https://www.jstor.org/stable/10.5325/jinfopoli.8.2018.0116

**Stalder F** (2010) Digital commons. In Hart K, Laville J-L and Cattani AD (eds), *The Human Economy. A Citizen's Guide*. Polity Press, pp. 313–324.

**Stalla-Bourdillon S**, **Carmichael L and Wintour A** (2021) Fostering trustworthy data sharing: establishing data foundations in practice. *Data & Policy 3*, e4. https://doi.org/10.1017/dap.2020.24

**Stalla-Bourdillon S**, **Thuermer G**, **Walker J**, **Carmichael L and Simperl E** (2020) Data protection by design: building the foundations of trustworthy data sharing. *Data & Policy 2*, e4. https://doi.org/10.1017/dap.2020.1

**Stalla-Bourdillon S**, **Wintour A and Carmichael L** (2019) Building trust through data foundations; a call for a data governance model to support trustworthy data sharing. Available at https://cdn.southampton.ac.uk/assets/imported/transforms/content-

https://doi.org/10.1017/dap.2021.40 Published online by Cambridge University Press

block/UsefulDownloads_Download/E2360AAB5D274223BFDB863BAFC20F34/White%20Paper%202.pdf (accessed 26 August 2021).

**Stilgoe J**, **Lock SJ and Wilsdon J** (2014) Why should we promote public engagement with science? *Public Understanding of Science 23*(1), 4–15. https://doi.org/10.1177/0963662513518154

**Strandburg KJ**, **Frischmann BM and Madison MJ** (2017) The knowledge commons framework. In Strandburg KJ, Frischmann BM and Madison MJ (eds), *Governing Medical Knowledge Commons*. Cambridge Studies on Governing Knowledge Commons, pp. 9–18. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316544587.002

**Susha I**, **Janssen M and Verhulst SG** (2017) Data collaboratives as "bazaars"? A review of coordination problems and mechanisms to match demand for data with supply. *Transforming Government: People, Process and Policy 11*. https://doi.org/10.1108/TG-01-2017-0007

**The Data Transfer Project** (2018) Data transfer project. Available at https://datatransferproject.dev/ (accessed 26 August 2021).

**The Global Partnership of Artificial Intelligence** (2021) Enabling data sharing for social benefit through data trusts. Available at https://gpai.ai/projects/data-governance/data-trusts/ (accessed 1 November 2021).

**The GPAI Data Governance Working Group** (2021) Understanding data trusts. Available at https://ceimia.org/wp-content/uploads/2021/07/2021-07-09-GPAI-summary-understanding-data-trusts-updated.docx.pdf (accessed 1 November 2021).

**Tironi M** (2015) Disastrous publics: counter-enactments in participatory experiments. *Science, Technology, & Human Values 40*(4), 564–587. https://doi.org/10.1177/0162243914560649

**Verhulst SG** (2021) Reimagining data responsibility: 10 new approaches toward a culture of trust in re-using data to address critical public needs. *Data & Policy 3*, e6. https://doi.org/10.1017/dap.2021.4

**Verhulst SG and Sangokoya D** (2015) Data collaboratives: exchanging data to improve people's lives. Available at https://sverhulst.medium.com/data-collaboratives-exchanging-data-to-improve-people-s-lives-d0fcfc1bdd9a (accessed 26 August 2021).

**Verhulst SG**, **Young A and Srinivasan P** (2021) An introduction to data collaboratives: creating public value by exchanging data. Available at https://datacollaboratives.org/static/files/data-collaboratives-intro.pdf (accessed 26 August 2021).

**Wikimedia** (2021) Commons Project Scope. Available at https://commons.wikimedia.org/wiki/Commons:Project%5C_scope (accessed 26 August 2021).

**Wilkinson MD**, **Dumontier M**, **Aalbersberg IJ**, **Appleton G**, **Axton M**, **Baak A**, **Blomberg N**, **Boiten JW**, **Bonino da Silva Santos L**, **Bourne PE**, **Bouwman J**, **Brookes AJ**, **Clark T**, **Crosas M**, **Dillo I**, **Dumon O**, **Edmunds S**, **Evelo CT**, **Finkers R**, **Gonzalez-Beltran A**, **Gray AJG**, **Groth P**, **Goble C**, **Grethe JS**, **Heringa J**, **Hoen PAC**, **Hooft R**, **Kuhn T**, **Kok R**, **Kok J**, **Lusher SJ**, **Martone ME**, **Mons A**, **Packer AL**, **Persson B**, **Rocca-Serra P**, **Roos M**, **Schaik R**, **Sansone SA**, **Schultes E**, **Sengstag T**, **Slater T**, **Strawn G**, **Swertz MA**, **Thompson M**, **Lei J**, **Mulligen E**, **Velterop J**, **Waagmeester A**, **Wittenburg P**, **Wolstencroft K**, **Zhao J**, **Mons B** (2016) The FAIR guiding principles for scientific data management and stewardship. *Scientific Data 3*(1). https://doi.org/10.1038/sdata.2016.18

**Wilsdon J and Willis R** (2004) *See-Through Science: Why Public Engagement Needs to Move Upstream*. London: Demos.

**Wilsdon J**, **Wynne B and Stilgoe J** (2005) *The Public Value of Science*. London: Demos.

**Wong J and Henderson T** (2019) The right to data portability in practice: exploring the implications of the technologically neutral GDPR. *International Data Privacy Law 9*(3), 173–191. https://doi.org/10.1093/idpl/ipz008. eprint available at https://academic.oup.com/idpl/article-pdf/9/3/173/31063862/ipz008.pdf

**Wong J and Henderson T** (2020) Co-creating autonomy: group data protection and individual self-determination within a data commons. In Ashley K, Whyte A, Pitkin K, Delipalta A and Sisu D (eds), *Proceedings of the 2020 International Data Curation Conference*. Dublin, Ireland: Data Curation Center.

**Wynne B** (2006) Public engagement as a means of restoring public trust in science—hitting the notes, but missing the music? *Community Genetics 9*(3), 211–220. Available at https://www.jstor.org/stable/26679532

**Young A and Verhulst SG** (2020) Data collaboratives. In Harris P, Bitonti A, Fleisher CS and Skorkjær Binderkrantz A (eds), *The Palgrave Encyclopedia of Interest Groups, Lobbying and Public Affairs*, pp. 1–5. https://doi.org/10.1007/978-3-030-13895-0_92-1

**Cite this article: Wong J, Henderson T and Ball K** (2022). Data protection for the common good: Developing a framework for a data protection-focused data commons. *Data & Policy*, 4: e3. doi:10.1017/dap.2021.40