Data Protection for the Common Good: Developing a framework for a data protection-focused data commons

Janis Wong*, Tristan Henderson, and Kirstie Ball

{jccw, tnhh, kirstie.ball} @st-andrews.ac.uk

Abstract

In our data-driven society, personal data 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 responsibilisation of data protection. To better protect individual autonomy over personal data, we posit that a data protection-focused data commons framework can be developed to encourage co-creating data protection solutions, rebalancing power between data subjects and controllers. Conducting interviews with commons experts, we aim to better understand how data protection were considered in existing commons and how privacy principles can be better applied. Incorporating trust, multidisciplinary knowledge, and public participation, a data protection-focused data commons can represent a community network of norms and values, enabling the protection of personal data by considering data protection for the common good.

Keywords— Data protection; privacy; personal data; commons; interviews.

1 Introduction

Rapid technological innovation has changed how we, as individuals, interact with companies using our personal data in our data-driven society. While data protection and privacy laws and technologies attempt to address concerns about data breaches and privacy scandals, they inadequately protect personal data (Kammourieh et al., 2017). Current approaches to data protection rely on a high-level of understanding of both the law and the resources available for individual redress, resulting in the responsibilisation of data protection (Mahieu, Asghari, & van Eeten, 2017). Frameworks such as data trusts and data collaboratives have been considered for protecting data subjects, but may not include them as part of the data protection process (Open Data Institute, 2019). Without direct data subject engagement, individuals and groups of data subjects may be excluded from participation where they are only the potential beneficiaries and are not part of designing the frameworks.

Using commons principles and theories (E. Ostrom, 1990), we suggest that a commons for data protection, a "data commons", can be created to allow 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 this paper we present empirical work conducted with nine interviewees from six data commons to examine how data protection can be best considered in a commons. We find that although the commons provides control and transparency of how data were collected, used, and processed, there were limited applications to wider data protection principles such as those about data subject rights. Interviewees further mentioned 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. This suggests that data subjects may be given greater control over their personal data by including them in the data protection process, while acknowledging knowledge gaps compared to experts or other stakeholders.

2 Background

Given the limited ability for data subjects to voice their concerns and participate in the data protection process, we posit that protecting data from harms resulting from mass data collection, processing, and sharing could be improved by involving data subjects in co-creation through a commons.

2.1 The Commons

Developed by Elinor Ostrom, the commons considers individual and group collective action, trust, and cooperation (E.

Ostrom, 1990). The framework 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 over-exploited. The commons depends on human activities and CPR management follows the norms and rules of the community autonomously (E. Ostrom, 1990), where "each stakeholder has an equal interest" (Hess, 2006). Central to governing the commons is recognising polycentricity, a complex form of governance with multiple centres of decision-making, each of which operates with some degree of autonomy (V. Ostrom, Tiebout, & Warren, 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, as illustrated by Ostrom's case studies of Nepalese irrigation systems, Indonesian fisheries, and Japanese mountains. These commons structures have enabled communities to find stable and effective ways to define CPR boundaries, define the rules for its use, and effectively enforce those rules (E. Ostrom, 2012).

Acknowledging 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 recognises that information is no longer a free and open public good and now needs to be managed, monitored, and protected for sustainability and preservation. In assessing the feasibility of a knowledge commons, Ostrom's Institutional Analysis and Development (IAD) framework is used to study an institution's community, resource dynamics, and stakeholder interests. 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 behaviours", analysing the "dynamic situations where individuals develop new norms, new rules, and new physical technologies" (Hess & 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 organisations, 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.

2.2 Urban commons and data commons for transparency and accountability

An urban commons represents resources in the city which are managed by the users in a non-profit oriented and prosocial way (Dellenbaugh-Losse, Zimmermann, & de Vries, 2020). It is a physical and digital environment that aims to better utilise an urban space for the public good, formed through a participatory, collaborative process. Data commons frameworks have been applied to urban environments in an attempt for governments to take more responsibility over its citizens' personal data (European Commission, 2018). With dynamic consent (Kaye et al., 2015), 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. Resource management "is characteristically oriented towards use within the community, rather than exchange in the market" (Stalder, 2010). An urban commons and its similarities to a digital commons are represented as information resources created and shared within voluntary communities.

Other data commons include those that focus on data distribution rather than data protection. Research data commons such as the Australia Research Data Commons (2020), the Genomic Data Commons (National Cancer Institute, 2020), and the European Open Science Cloud (European Commission, 2019) all attempt to further open science and open access initiatives. While these frameworks recognise that the information and knowledge are collectively created, their implementations are hierarchical and top-down without input from archive participants or repository managers. Additionally, existing commons frameworks do not protect the personal data within them as they prioritise data sharing over data protection, particularly on data curation and reuse. As a result, we focus our work on looking at data commons applied to cities and urban commons.

2.3 Research questions

In previous work, we developed a data protection-focused data commons (Wong & Henderson, 2020): Figure 1 shows how a data subject specifies to what extent they would like their data to be protected based on existing conflicts preidentified within the data commons for a specific use case. In this study, we conduct interviews with commons experts to identify the challenges of building a commons and important considerations for a commons' success.

We established four research questions to explore whether using information rights 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 be best 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?

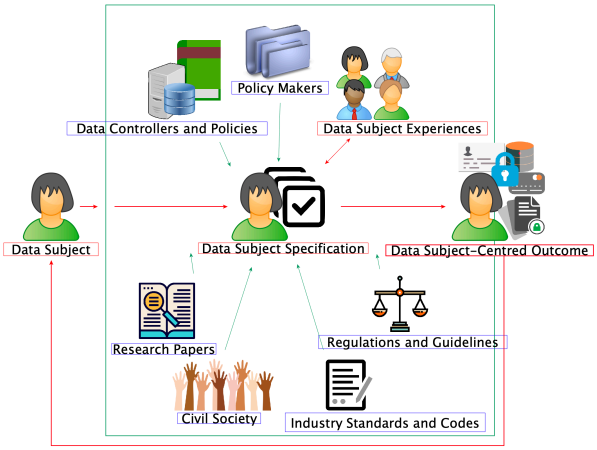

Figure 1: In a data protection-focused data commons (green), the data subject specifies to what extent they would like their data to be protected based on existing conflicts and challenges pre-identified within the data commons for the use case (red). No prior knowledge of existing law, norms, or policies are required. Along with stakeholder information (blue), the data subject specification is then used to inform their data protection outcome that is generated from the system. As the outcome is data subject-centred, decisions ensuring the protection of the data subject's personal data may override existing preferences, policies, or standards set by other stakeholders. Data subjects can return to and review their outcome, add their data subject experiences to the data commons, and participate in the co-creation process at any time.

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 it represents 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 the recent urban commons and data commons. As all authors reside within the jurisdiction of the EU General Data Protection Regulation (GDPR), an online search was conducted to identify European commons.

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. Interviews were conducted to contextualise 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.

3.2 Writing the interview questions

The interview questions reflect those of key informant interviews and were in a semi-structured format to encourage discussion around the commons. The questions aimed to answer the research questions identified in Section 2.3, 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.

3.3 Conducting the interviews

Interviews were conducted either over the phone or on 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 July 2020 and lasted within one hour. All interviews were recorded with the interviewee's consent. Once each interview was completed, the audio file was put into the MaxQDA qualitative data analysis software where it was manually transcribed and pseudonymised as soon as possible.

4 Analysis

Nine experts across six commons were interviewed. The size, number of participants, and stakeholders varied across the commons, with 3 interviewees based in The Netherlands, 2 in the United Kingdom, 1 in Belgium, 1 in Germany, 1 in Italy, and 1 in Spain. Their roles and specialism are listed in Table 1. The Reference *Cx* denotes the commons they contributed to and *Ex* denotes the expert. Individual roles are

characterised based on their commons-related work. Expertise describes their main contribution towards the commons.

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

Table 1: Reference of interviewees representing their commons project, role within the project, and their expertise.

We found that data protection within existing commons frameworks was predominantly considered only in terms of control and sovereignty of 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. Interviewees further mentioned 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

First, in identifying the data protection challenges within commons projects, interviewees mentioned that the main problems were provided by the project coordinators, with partial input from the experts themselves. One interviewee said: *"The project was, we have these technologies, we do not know how these are going to be because we haven't built it yet but it will revolve around data sovereignty and it is going to have state of the art technology"* (C2E1). An interviewee in a technical role added that following their core commons aim: *"The most important challenge there was to make it decentralised"* (C1E1). Regarding data protection, one interviewee elaborated that the scenario was created for those who worked on the project: *"Essentially what [the coordinators] wanted was, they realised that this poses a threat to [users'] privacy and they wanted us to build a system from the same dataset"* (C2E2).

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.

4.2 Overcoming data protection challenges

According to the experts, user trust in both the commons framework and those who created the framework was important for the common's 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 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 grounds, 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 doesn't 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). Regardless of the use case of the commons, it is important to understand the issues of the community when getting them involved, applied both to data protection and other issues. Interviewee C5E1 explained that although the community want to engage, they either do not know how or the way there were approached did not interest them. In the context of involving the community in data protection conversations, this could include knowing what their data related worries are, what issues data subjects are currently facing, and supporting their data protection rights.

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), emphasising the importance of timely development. Even in a commons, other stakeholders may be prioritised over data subjects: *"We played a role of responsibility, coordination, and interaction with [data protection officers] rather than data subjects"* (C3E2). Without 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 decision based on what the community want or what the stakeholders want or what the users want when the people we are hearing from are entirely unrepresentative of the local community"* (C5E1).

However, 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: *"The first thing I became aware of is the inequality in our access to the internet. 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.

4.4 Building a commons for data protection

One key point reiterated by many experts was the transition between theory and practice: *"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 finding for but at the same time or seeing to the sovereignty of citizen. It is one thing to understand what does this look like but in practice, how can we operationalise this?"* (C2E3). Several interviewees mentioned that as using a data commons is a choice, its purpose needs to be clear. 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, how it works, [and] how [data] has been managed" (C3E2). Importantly, individuals and communities need to be actively encouraged and empowered to co-create: *"A lot of people do commoning but they don't know they are commoning. They don't have an identity that permits them to have, to exert directly power"* (C3E1).

5 Discussion and Future Work

From our interview findings, we are adapting existing theories on the knowledge commons framework to develop a data protection-focused data commons. This includes applying the IAD framework with data protection information to create a commons and analyse the dynamic situations where individuals develop new norms, rules, and physical technologies to study the commons' community, resource dynamics, and stakeholder interests.

5.1 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 relevance of online teaching and remote learning, we are now creating data commons tools to be tested for this particular use case. 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, 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.

5.2 Data commons policies

Policies for creating a data protection-focused data commons can also be established to support the implementation of a commons. Establishing the use case domain and requirements such as listing the stakeholders involved, rules and norms of participation, and how the data protection artefacts within a commons should be protected, the creator of the commons, whether it may be an individual or an organisation, can map out the necessary requirements for a data commons. Using existing data protection policies, such as regulations and institutional policies or codes, as well as writing new community policies can support data subjects to co-create 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. For data controllers, this could be useful to better understand what data protection requirements are preferred by data subjects. Additionally, 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.

6 Conclusion

In this paper, we set out how a collaborative and co-created data protection-focused data commons will 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. Adopting existing commons frameworks and interviewing experts to learn about how data protection was considered in the commons and how challenges in the development process were overcome, data protection can be improved as a common good.

References

Australia Research Data Commons. (2020, July 17). Australia research data commons. Retrieved July 17, 2020, from https://ardc.edu.au

Dellenbaugh-Losse, M., Zimmermann, N.-E., & de Vries, N. (2020). *The urban commons cookbook: Strategies and insights for creating and maintaining urban commons*. La Vergne, Tennessee: IngramSpark.

European Commission. (2018). Reclaiming the smart city: Personal data, , and the new commons. *Decode*. Retrieved from 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*. Retrieved from https://op.europa.eu/en/publication-detail/-/publication/78ae5276-ae8e-11e9-9d01-01aa75ed71a1/language-en

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

Hess, C., & Ostrom, E. (2007). *Understanding knowledge as a commons: From theory to practice*. Cambridge, Massachusetts: MIT Press.

Kammourieh, L., Baar, T., Berens, J., Letouzé, E., Manske, J., Palmer, J., ... Vinck, P. (2017). Group privacy in the age of big data. In L. Taylor, L. Floridi, & B. van der Sloot (Eds.), *Group privacy: New challenges of data technologies"* (pp. 37–66). doi:10.1007/978-3-319-46608-8_3

Kaye, J., Whitley, E. A., Lund, D., Morrison, M., Teare, H., & Melham, K. (2015). Dynamic consent: A patient interface for twenty-first century research networks. *European Journal of Human Genetics volume 23, pages141–146(2015)*, *23*, 141–146. doi:10.1038/ejhg.2014.71

Mahieu, R., Asghari, H., & van Eeten, M. (2017). Collectively exercising the right of access: Individual effort, societal effect. In *Giganet (global internet governance academic network) annual symposium 2017*. doi:10.2139/ssrn.3107292

National Cancer Institute. (2020, July 17). Genomic data commons. Retrieved July 17, 2020, from https://gdc.cancer.gov/

Open Data Institute. (2019, April 15). Data trusts: Lessons from three pilots. Retrieved July 17, 2020, from https://theodi.org/article/odi-data-trusts-report/

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge, UK: Cambridge University Press.

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

Ostrom, E. (2012). *The future of the commons: Beyond market failure & government regulations*. London, UK: Institute of Economic Affairs.

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

Stalder, F. (2010). Digital commons. In K. Hart, J.-L. Laville, & A. D. Cattani (Eds.), *The human economy. a citizen's guide* (pp. 313–324). Polity Press.

Wong, J., & Henderson, T. (2020). Co-creating autonomy: Group data protection and individual self-determination within a data commons. In *Proceedings of the 15th international digital curation conference*.