Trade-offs of data collection and digitalization in cooperatives

A study in the context of a low digital literacy setting

Master's Thesis
Hilja Tukia
Aalto University School of Business
Global Management
Spring 2023

Author	Hilja Tukia
Title of thesis	Trade-offs of data collection and digitalization in cooperatives
Degree	Master of Science in Economics and Business Administration
Degree programme	Global Management
Thesis advisor(s)	Patrick Shulist
Year of approval 2023	**Number of pages** 75 **Language** English

Abstract

In an increasingly digitalizing world, the importance of data is becoming progressively more evident. With data even being described by some as the 'new oil', there is a rising importance placed on data. Through this new value attached to data, comes along a multitude of new questions and considerations. As with digitalization, not everyone benefits equally from datafication. The reasoning behind how and why data is being collected becomes increasingly more critical as data is used to make important decisions – thus creating the acute need for proper data governance.

Cooperatives are member-driven organizations where decisions are made democratically. The purpose of cooperatives is to enhance members lives, economically, socially, and culturally (International Cooperative Alliance, 2023). Cooperatives are seen as an alternative to more traditional corporations, thus with the increasing value of data the need to cooperatives to diversify into the the digital sphere emerges.

In this thesis, trade-offs that cooperatives operating in settings with low digital literacy make when digitalizing and using data are explored. The thesis aims to reflect on the implications of the different trade-offs identified through both existing literature as well as semi-structured interviews conducted with individuals working at cooperative federations.

The digital divide affects ability to participate in many digital and thus data-related activities, however digitalization of cooperatives is seen as a solution to this. Although benefits of cooperatives using data were many, risks highlighted were primarily related around informed consent, lack of resources, relevance of the data selected for collection, the burden of time taken, and the security of the data itself. Meanwhile, benefits identified were greater access to information, increased sense of agency, potential for entrepreneurship in the digital sphere, improved decision-making, and potential to learn and educate more about the benefits of data and digitalization.

Keywords digital divide, cooperatives, cooperative federations, data, data governance

Table of Contents

LIST OF FIGURES

0 Definitions

To begin the thesis, I would like to provide some key definitions for readability purposes.

Artificial Intelligence (AI) – AI in this thesis is defined as in Encyclopedia Britannica as "the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings." (Copeland 2021). Algorithms are in turn the processes which allow AI to function.

Data – There is a distinction between data and information (Weber et al. 2009), as data is simply a point of information that without context does not contain any information (Zins 2007, p.481). Thus, for the purpose of this thesis, when the term data is used it should be assumed that this data is informative.

Information Communication Technology (ICT) – ICT is digital tools such as mobile phones and laptops.

1 Introduction

High market share of large corporations, digital exclusion, unsatisfactory data governance, skill and capacity gaps, and diminishing data sovereignty are all seen as being the major challenges towards data usage in the future (Bühler et al. 2023, p.148). The value of data is vast, with data-driven firms leading in market capitalization in 2021 (Coyle & Manley, 2022). Data is at the crux of digital usage and used to develop new technologies, create value for companies, inform decision makers, and many other applications. The digital divide also highlights other inequalities such as the data divide. Both outcomes are examined in books such as *the Invisible Women,* where Perez (2019) highlights the different experienced outcomes for women when considering day-to-day situations due to differences in data biases involved in the design of these systems.

Information communication technology (ICT) can be seen as a tool for development, and a way to decrease inequalities globally (Brewer et al. 2005). However, the outcomes of using ICT for development must be looked at further through different cases. One of the reasons for this is the digital divide. The digital divide is a global phenomenon which is defined as "a division between people who have access and use of digital media and those who do not" (Van Dijk 2019, p.1).

Data governance and artificial intelligence "AI" is a field growing in prominence as technology develops. Data, referred to some as "the new oil" (N.A, N.D, Crawford 2021, p. 113) is seen to create value (Crawford 2021). There are many ways in which data used for AI can be collected and henceforth governed, however with these methods there are also risks in part. Data governance focuses mainly on the management of data; however, this does not always include the full data lifecycle as data collection and quality of this collection are not seen to be part of this lifecycle of data.

Mistakes in data governance can have profound effects, that may lead to outcomes such as AI perpetuating issues such as systemic biases, unlawful decision making, financial losses, political crises, and even loss of life (Janssen et al. 2020, p.2). Additionally, even just the methodology, source and the 'coders' of data for AI may unintentionally play a role in creating the aforementioned issues in an AI system (Crawford 2021), which relied on good

data governance for accuracy. Data governance risks are not limited to biases in data sets leading to undesirable outcomes, but improper data governance can lead to data security risks. In India, data governance is an emerging matter. The Indian government released the Digital Personal Data Protection Bill in 2022, replacing the Personal Data Protection Bill from 2019 (MEITY 2022; KPMG 2022). The new bill introduces requirements for explicit consent in collecting data and presents writes for the data principals – which are the people, organizations or other entities providing the data (MEITY 2022; KPMG 2022).

Cooperatives are member-governed organizations that exist in many fields including producer, financial, and worker. In recent years the cooperative movement has involved to better serve members. The emergence of platform cooperatives, data cooperatives, and other forms of digital cooperatives represent the growing evolution of cooperatives to better serve their members and communities. Cooperatives in general can be seen to confront capitalistic tendencies (Ranis 2016). Data cooperatives provide a way to distribute and utilize data amongst members, as members continue to own the rights to this data. This contrasts with corporate data governance solutions, where the company is the one profiting from the data (Crawford 2021). Data cooperatives build on the understanding that data and data governance are collective rather than private property (Miller 2021). As mentioned before, data is valuable and is becoming increasingly commodified. Accordingly, the need for cooperatives to diversify into data and digital spaces arises.

Research Objective and Questions

With the rising need to implement data governance solutions and ensure that individuals are able to maintain rights to and benefit from their own data, the situation of data in cooperatives should be further examined to understand the risks and benefits, especially in the case of contexts where there is low digital literacy – meaning that the value of own personal data may not be fully understood. Due to the intrinsic linkage between data and use of ICT, outcomes of cooperatives digitalizing should be studied alongside data as the two are interlinked.

In order to better explore the topic of the digital divide, data governance and how cooperatives fit in. I have selected three research questions.

Firstly, to get a better understanding of the context of the case I will be studying my first research question is as follows:

RQ1: What are the trade-offs of collecting data in cooperatives?

This question aims to explore benefits and risks posed by data collection and digitalization in cooperatives. As cooperatives further digitalize, data can be applied towards new solutions. This means that it is important to consider the type of risks and benefits that are invited intentionally or unintentionally through collecting and processing data in the cooperative setting.

Due to the specifics of the case, I have selected two more research questions to assist in further developing the topic:

RQ2: What are the digital divides faced by cooperatives?

The aim of this question is to identify the types of digital divides faced by cooperatives. This is specifically relevant as the case chosen for the thesis explores data collection and use of ICT in settings where the digital divide is present. The digital divide can potentially create differences in how the outcomes of data collection are seen in cooperatives, and since the aim of data cooperatives is to create value for members (Hardjono & Pentland 2019, p.2). Thus, I wish to understand what kind of digital divides are faced to evaluate the kind of trade-off the digital divide can have towards data collection. Additionally, Bühler et al. (2023) list digital exclusion and skill and capacity gaps as being challenges data cooperatives are faced with. Hence, I will be examining what kinds of digital divides are encountered in the case selected.

RQ3: How is data governance in cooperatives affected by the digital divide?

Finally, the aim of this third and final question is to apply the principles of identifying the digital divide towards evaluating the impacts there may be towards data governance. As outlined above, good data governance is needed for data to be collected, stored, and used securely. However, the aim is to explore what happens to data in cases where digital and data literacy is low.

Thesis Structure

After the introductory chapter which briefly introduces the reader to the topic and objectives of my thesis, the thesis takes on a structure as such:

Firstly, the thesis begins with the literature review where I both provide theoretical background to the thesis. Here topics central towards the research objective and the research question are presented. These include: background on cooperatives, as well as challenges they may face; digital and data divides; possible outcomes of ICT usage; and data governance frameworks. The literature review will serve as the basis on which the research gap of my thesis is positioned.

Secondly, the selected methodology for the thesis will be presented. Here the methodology of the thesis from how the research was conducted, to the analysis of the data collected will be explained. Additionally, justification for selecting the chosen methodology will be provided.

Thirdly, findings are organized through a framework demonstrating where different presented codes identified during the analysis are present. This will be followed by a discussion section, where findings will relate to existing literature – and different points in the findings will be further discussed. Additionally, the discussion section will provide potential implications of the research.

Finally, the thesis concludes with a summary of the research in which the research questions are answered. In the conclusion, I will also present some of the limitations of my thesis, as well as make some suggestions for interesting topics for future research.

2 Literature Review

In this section of the thesis, the aim is to provide some theoretical background about the topic and build toward the research gap based on existing literature. This will be done in four sections: cooperatives, digital divide, outcomes of ICT and Data Use, Data Governance.

Cooperatives

The purpose of this section is to introduce cooperatives and the principles they are built on, the purpose is also to provide some information about participation and inclusivity of cooperatives.

Fundamentals of Cooperatives (Background)

According to the International Cooperative Alliance (ICA) cooperatives are defined as:

> "*an autonomous association 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 2023)

Cooperatives are organizations which members join voluntarily to collectively organize to pursue a joint goal. Cooperatives are governed democratically, with the one member one vote rule (International Cooperative Alliance 2023). Cooperatives are both governed and owned by the members, meaning that cooperative members may find more autonomy in their economic and social livelihoods through membership. There are seven cooperative principles listed by the International Cooperative Alliance (2023), these are:

> "1. *Voluntary and Open Membership*, 2. *Democratic Member Control*, 3. *Member Economic Participation,*, 4. *Autonomy and Independence*, 5.*Education, Training and Information*, 6. *Cooperation among Cooperatives*, 7. *Concern for the Community*" (International Cooperative Alliance 2023)

Puusa et al. (2013) found that cooperatives are viewed as producing benefits for both members and society, and the decision-making democracy in cooperatives was viewed positively, which aligns with the seven cooperative principles. However, Puusa et al. (2013)

also found that the cooperatives are considered inflexible in comparison to corporations, due to bureaucracy. This is quite natural, when taking into consideration the fact that every member in the cooperative has a vote which creates more bureaucracy, while in other governing mechanisms decisions can be made more unilaterally without requiring a vote.

Ranis (2016, p. 40) writes "Cooperative formation shows the demands of workers in all forms of political systems and aspires thus to greater democratic participation in their lives and enhanced worker self-management in their lives". This means involvement of workers involved in cooperatives on a civil society level. The implication for data cooperatives could mean much of the same, involvement and greater participation to produce results favorable to workers themselves. The concepts of cooperatives are important towards answering the research questions as the case study of the thesis will be based on cooperatives. The idea of recuperation is also fitting towards data, as data is held in the hands of few (Hardjono & Pentland 2019) and exploited for profit to those who 'own' the data – regardless of if this data is held consensually (Crawford 2021).

Cooperatives in India

With the state withdrawing from agricultural markets, India faces new challenges and opportunities in structuring the agricultural market (Singh 2012). The agricultural market in India is comprised mainly of smallholder farmers (ibid). This means that with government withdrawal from the agricultural market, smallholders must be finding different ways in which to support their livelihood. One of the problems smallholders' faces is exclusion from the market, for example, smallholders lack bargaining power and thus receiving lower prices on their products than large-scale farmers (Singh 2012). Additionally, barriers regarding quality control and delayed payments until inspection of products can prevent smallholders from accessing certain markets (ibid.). Thus, cooperatives are a way in which smallholder farmers can collectively negotiate better prices (Singh 2012). In addition to economic incentives, cooperatives may also provide smallholders with opportunities to receive professional training in marketing and production technologies (Singh 2012, pp. 101, 104). With the support of cooperatives farmers can collectively negotiate for better prices on their crops, as well as gain education on new farming techniques and improving the quality of their crops.

Inclusivity and Participation in Cooperatives

Despite the intentions of cooperatives to promote economic and social well-being, as well as enhance democratic participation – sometimes in practice divides within cooperative members can lead to challenges in inclusivity and participation within cooperatives. In this section, the participation and inclusion in cooperatives as well as potential challenges towards it will be explored.

Ito et al. (2012, p. 707) find that small scale farmers earn double the benefit from being part of a cooperative, in comparison to large scale farmers. Empirical evidence collected from cooperative member and nonmember smallholder dairy farmers in Bihar also points to higher income gains for smallholder cooperative members (Kumar et al. 2018). This demonstrates a clear benefit of increased income cooperatives for small scale farmers who are cooperative members. Ito et al. (2012, p.708) argue that including smallholders is propitious in alleviating rural poverty, while also being a deterrent of profitability of cooperatives. Thus, smallholders do benefit from membership in cooperatives, however depending on the market-orientation of the cooperative it may be that they are not even accepted into the cooperative, despite cooperative principles demanding open membership (International Cooperative Alliance, 2023). Bijman & Wijers (2019) reaffirm this by presenting a framework in which producer cooperatives that are more market-oriented than community-oriented will be more exclusionary to those who do not own land or have a lower income. Bijman & Wijers (2019) describe cooperatives that are more "community-oriented" as ones which implement social activities at an expense of economic performance, and those that are "market-oriented" as cooperatives which take a more strategic economic approach and focuses on economic motivations over social ones. Bijman & Wijers (2019) reiterate the finding from Ito et al. (2012), that there are economic arguments to exclude some members, for example if the farmers cannot contribute to economic efficiency or fit the cooperatives strategy due to reasons such as not being able to meet produce quality requirements of a cooperative (which can provide more bargaining power to cooperatives if can be standardized (Singh 2012).

In addition to member acceptance in cooperatives being impacted by the orientation of the cooperative, the inclusion and representation of members within the cooperative may be impacted by socio-economic difference between cooperative members. Exclusion of smallholder farmers in cooperative governance may exist due to factors such as remote

location of farm and low levels of education (Bijman & Wijers 2019). There are also transaction costs to participation, which for example could be if a smallholder is remotely located, it could be costly to travel to cooperative meetings where governance decisions may be made. Additionally, sometimes due to social reasons woman are excluded from the cooperative decision-making process (Bijman & Wijers 2019). Meaning that inclusivity of a cooperative shouldn't necessarily be assumed, as there are factors that may bar participation even if one is permitted to join a cooperative. As members are not always represented equally in cooperative governance, it is important to study what factors facilitate inclusion in cooperatives. However, agricultural cooperatives are still seen as being part of a strategy enhancing women's economic and social empowerment (Dohmwirth &Hanisch, 2019).

Although democratic voting of members is seen as part of the cooperative principles, it is essential not to overlook the fact that some members may still face barriers especially if the transaction costs to travel to cast a vote are high – meaning that members with better means of transportation or less-rurally located may be able to influence cooperative activities better into their own favor through voting accordingly.

Figure 1: Characteristics of producer cooperatives from Bijman & Wijers (2019, p.77)

Key characteristics of the cooperative	Rural development focus	Development and business focus	Business focus
Main values	Solidarity	Solidarity and Efficiency	Efficiency
Orientation	Community	Community and Market	Market
Membership	Open	Semi-closed	Closed
Inclusiveness	High	Medium	Low

In Figure 1, a matrix from Bijman & Wijers (2019, p.77) describes characteristics of rural development focused, development and business focused and business focused cooperatives. The matrix describes business focused cooperatives as valuing efficiency (meaning efficiency in the operation of the cooperative) and thus the orientation of the cooperative lies towards market, and membership is closed to maintain the efficiency of the cooperative. This makes the inclusivity of the cooperative low, as members are excluded on, for example, economic basis for the cooperative to be market efficient. Alternatively, cooperatives with a "rural development focus" value solidarity, have a community orientation, open membership and are highly inclusive for members (Bijman & Wijers 2019). Bijman & Wijers (2019) also

suggest that a cooperative may follow a transition from a rural development focus to a business focused cooperative at some point, indicating that the role of cooperatives is not static but rather mobile.

Participation in cooperatives is not solely determined by the inclusivity of cooperatives. Cooperatives that have large membership bases may face a degrading quality in the social environment and moral involvement of its members in cooperative activities (Puusa et al. 2013). This may be due to challenges in member control in large heterogenous cooperatives (Österberg & Nilsson 2009), an example of this is because large cooperatives often serve many interests and have various branches. However, sometimes members are only interested in some branches of service provided by the cooperative, creating a divide in usage of cooperative services. Large-scale cooperatives may also create passivity in members due to individual voices not being heard due to the large scale of the cooperative, not all members will get all their interests attended to (Österberg & Nilsson 2009).

Österberg & Nilsson (2009, p. 191) found that the most committed cooperative members were those satisfied with the profitability of their own farm were most committed to their cooperative. However, these members had the lowest trust in the board of directors of the cooperative (Österberg & Nilsson 2009, p. 191). Trust in the board of directors is not directly linked to profitability, perhaps due to cooperatives having social aims and not being purely economic institutions. Social development aims of cooperatives thus remain important; however, it may be that more economically oriented members do not find these programs rewarding to their membership. However, as Österberg and Nilsson (2009) found that do not believe that cooperatives contribute to their farm profitability in this case, leading to the fact that social factors may be important in cooperative operations. Democratic governance of the cooperative is considered crucial (Österberg and Nilsson, 2009), thus democratic governance structures and fair governance structures should be prioritized in all aspects of a cooperatives. It is crucial to note that Österberg and Nilsson's (2009) research takes place in the context of Sweden, meaning that participation patterns may differ in other locations or cultural contexts.

Concluding Cooperatives Section

In this section, it is indicated that sometimes there is a gap between the theory (International Cooperative Alliances seven cooperative principles) and practice (inclusion and participation

of members in cooperatives). For the context of this thesis, this indicates that divides and differences in characteristics of members may lead to differing results in the ability to participate and benefit from the usage of data and digital tools to achieve desired outcomes by cooperatives.
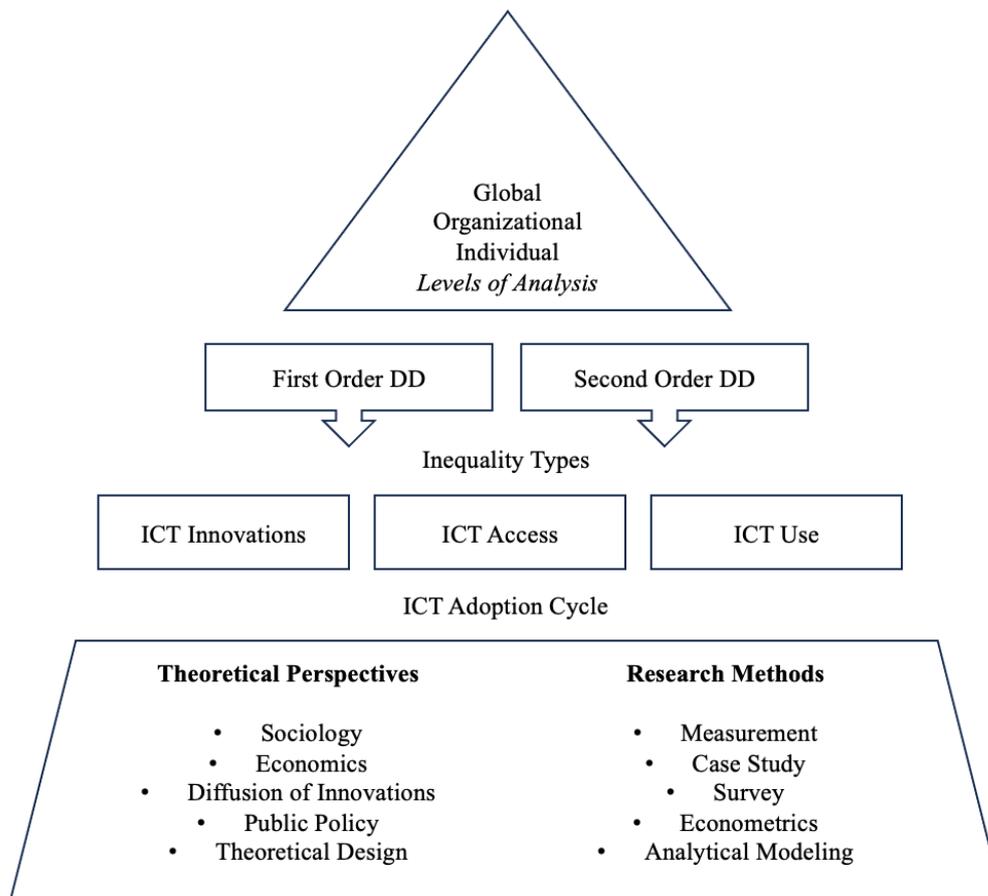
Digital Divide

In this section of the literature review I review literature related to the digital divide I begin by defining characteristics of the digital divide, I then introduce different frameworks. used to characterize aspects of the digital divide. The digital divide is one of the most relevant concepts to explore in response to my research questions as the phenomenon exists on different levels that shall be explored to give a more nuanced answer to the research questions.

Van Dijk (2019, p. 1) defines the digital divide as "a division between people who have access and use of digital media and those who do not". Digital media includes forms of technology such as computers, smartphones and other types of information and communication technology (ibid.) Although the digital divide is a widely studied phenomena, there are challenges towards defining the divide as "there is a lack of consistency in the terminology used, both for the type of digital divide addressed (skills, uses and outcomes), as well as for the determinants" (Scheerder et al. 2017, p.1608). This creates a situation in which it is uncertain which aspects of the digital divide are being discussed in certain studies, creating a situation in which the impact of the digital divide is difficult to place due to uncertainty of the origin. This means that unless otherwise specified, evidence of digital divides may be difficult to categorize as it is not always specified which type of digital divide is being referenced in literature.

As mentioned before by Scheerder et al. (2017), the digital divide has evolved from simply being characterized as a divide between people who have access and capability of using ICT, to one with further implications. Riggins & Dewan (2005, p.301) claim that diffusion of different ICT innovations is central to the digital divide, as these technologies are adapted at different rates with different capabilities leading to technological comparative advantages.

Figure 2: Conceptual Framework for Organizing Research on the Digital Divide (FRiggins & Dewan, 2005 p.302)


In Figure 2, Riggins and Dewan (2005, p.302) conceptualize how to organize research on the digital divide. From this framework, the illustration of the emergence of two types of digital divide types at different phases of the ICT cycle is important, as it demonstrates degrees of differences within the digital divide in relation to access and usage throughout the ICT Adoption Cycle. Riggins and Dewan (2005) also paired with the concept of these divide types having to be considered from three different stakeholder groups as seen in Figure 2. The reason for examining the digital divide from the different perspectives is because of the different degrees of effects and solutions the digital divide may have on these stakeholders.

Based on the existing framework by Riggins and Dewan (2005), Wei et al. (2011) identified three different levels of the digital divide. The three levels of digital divide in Wei et al.'s (2011) framework can be seen below in Figure 3. In Figure 3, one can see that the digital divide can be studied through the individual, organizational and global lenses, and the divide moves through different phases as concerns are addressed, and digital access and capability evolves. Wei et al. (2011, p. 182) find that the digital access divide influences the digital

capability divide, which further influences the digital outcome divide. The three levels (individual, organization, and global) of the digital divide are significant, as they expand the understanding of the phenomena to a broader scale and allow for deeper examination of implications of the digital divide, and the addition of formally acknowledging the outcomes divide withing the digital divide. The framework raises the question of whether digital access and capabilities necessarily lead to positive outcomes and demonstrate that inequalities between digital users may persist regardless of access and capability, as the outcomes may be less favorable to some.

Figure 3: Three- Level Digital Divide Framework (from Wei et al. (2011, p.171)

	IT Adoption Stages				
Individual Organization Global	Digital access divide	⇨	Digital capability divide	⇨	Digital outcome divide
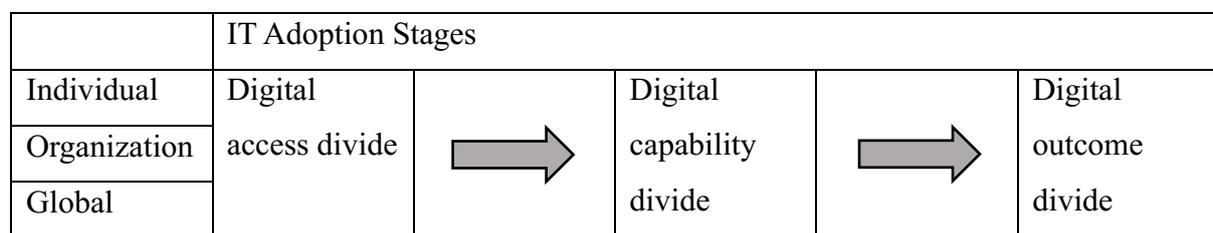
The first level of the digital divide is characterized as being the inequalities in access to ICT, whether this means lack of access to a personal device or a shared one, or to the internet in general (Riggins & Dewan 2005; Wei et al. 2011; Carter et al. 2020). The definition of the access divide varies across literature, for example Singh (2010) measures the first level or access divide in India through teledensity (number of landlines per 100 people), internet users (who have connected to the internet least once in the last month), and mobile users. However, these kinds of inquiries do not tell the full story. Singh (2010) also evaluates electrification rates as this is a precondition, thus the access divide can stretch deep.

The second level of the digital divide refers to inequalities in abilities and capabilities to use ICT amongst those who have access to ICT (Riggins and Dewan 2005; Wei et al. 2011; Lythreatis et al.. 2021). Despite having access to ICT, lack of digital literacy can cause digital division between people who are able to use technology such as mobile phones and computers, versus those who are not able to do so, or are less skilled at doing so. Age and gender are identified as key determinants linked with the first and second level digital divides, with gender being particularly linked to the first level divide of access. (Lythreatis et al. 2021). Van Dijk (2005) further depicts the access and capability divides as follows, sequentially:

1. **Motivational Access**

This is the motivation of the user to "adopt, acquire, learn, and use these technologies". (Van Dijk 2005, p. 27). The motivation of a user to continue using ICT is not stagnant, and user populations may be shifting in and out of using ICT. Lack of motivation for accessing ICT can range from various reasons such as lack of finances to purchase and fear of technology. (Van Dijk 2005).

2. **Material Access**

Material access if the physical access to ICT. This can occur through household ownership of a device such as a computer or smartphone, or even mean access to a computer at an internet café or library. However, the distinction of what constitutes as access and poor data on the subject inhibit studies of material access. (Van Dijk 2005) According to Van Deursen & Van Dijk (2019) some of the differences in material access are:

" (1) differences in device opportunities, or the use replacement of a device by other devices with different technical capacities; (2) differences in the diversity of devices and peripherals; and (3) differences in the maintenance costs of devices and peripherals." (Van Deursen & Van Dijk 2019, p.356).

3. **Skills Access**

Skills access is known as including operational, informational, and strategic technological skills. These are three different levels of skills one can have when using ICT. Operational can mean something like knowing how to use a computer and the applications, informational skills are knowing how to search for information on the computer, and finally strategic is knowing how to use a computer to better one's societal standing (Van Dijk 2005, p. 75).

4. **Usage Access**

The usage divide is a digital divide which occurs when all the three conditions from before are met, but the user still does not use the ICT or only uses it for limited hours. Things that can influence this are functionality, expense etc. (Van Dijk 2005)

As can be inferred, even though these steps are sequential – many of them are interlinked and thus it can be very difficult to measure the effects on their own.

The third level of the digital divide, defined by Wei et al. (2011, p. 171) is the "digital outcome divide". The digital outcome divide refers to the outcomes that arrive after using technology. The content of the outcome can vary depending on what is being studied, for example in the instance of Wei et al. (2011, p. 183) the outcome divide refers to the

difference in learning outcome gap between students who are using IT-based learning methods. Fernandez et al. (2020) discusses an outcome gap wherein people highly dependent on mobile phones as a source of technology are less likely to use the internet to combat socioeconomic divides (as cited by Lythreatis et al. 2021). Thus, the outcome gap introduces a situation in which the outcomes of using technology are vastly different for the end-user.

Figure 4: Causal model of resources and appropriation theory (From Van Dijk 2005, p.15)
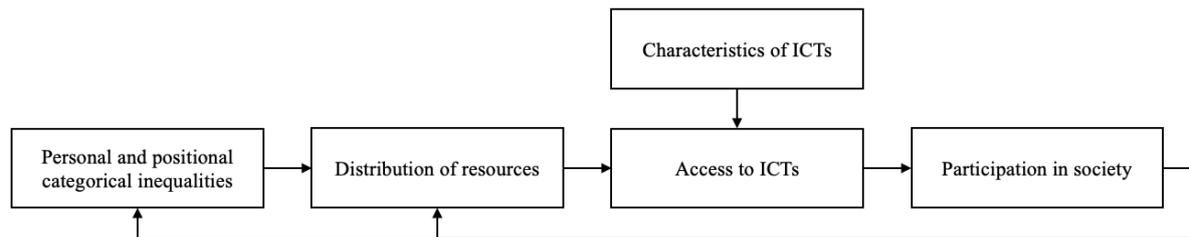
Van Dijk's characterization of the digital divide is slightly different to that of Wei et al. (2011) of the divide (Van Dijk 2005), however the two are not mutually exclusive as they share similar characteristics. Van Dijk explains the digital divide as a process of understanding different kinds of access, the causes for differences, and finally the outcomes or consequences of this (Van Dijk 2005, p. 14). Thus, while Wei et al. (2011) study the digital divide as more of a process, Van Dijk (2005) aims to find some kind of understanding of the digital divide as a part of a more extended system. Van Dijk's (2005, p.15) "causal model of resources and appropriation theory" as can be seen in figure 4 shows the relationship of how the digital divide feeds into societal inequalities. Participation in society can lessen inequalities and reinforce distributions of resources (Van Dijk 2005). However, inequalities also carry through the system as if ICT resources are distributed unequally, participation in society is diminished in this model (Van Dijk 2005). Additionally, Van Dijk acknowledges difference characteristics of ICTs can affect the level of accessibility to IT.

Lythreatis et al. (2021) suggests the digital divide stems not just across genders but also through other sociodemographic elements such as age, race, urbanization etc., as well as through other elements such as rights and social support. Gender is a way in which the digital divide in India. A difference in access to technology, differences in levels of digital literacy exists between men and women in India (Singh 2010). Singh (2010) finds that woman have less access to technology and lower digital literacy levels (Singh 2010). This creates a situation in which the digital divide disproportionately effects woman, in comparison to men. As digitalization increases, more social divides are realized– as is theorized in Figure 4 – this

18

can include inequalities in access and digital literacy between genders. Rajam et al. (2021), find that there are large digital divides between some caste groups in India in comparison to others. The digital divides referred to here are the access and the digital literacy level divides (ibid.). The digital divide can exacerbate social divides due to the nature of digitalization. Research by Tewathia et al. (2020) seem to agree that digital divides perpetuate social inequalities, and that ICT amplifies inequalities in India. Tewathia et al. (2020) and Jamil (2020) suggest using ICT to enhance digital inclusion rather than just taking actions to reduce the digital divide. Tewathia et al. (2020), note that there is limited academic and policy literature focusing on causes and effects of the digital divide in India.

Data Divide

The digital and the data divide have many similarities between the outcomes and the reasoning behind the existence of both divides. Shah et al. (2022, p.432) describe the data divide as "gaps in the ability of individuals and communities to collect, access, integrate, store, analyze, and use data to create value, including commercial value." This is not dissimilar to the digital divide wherein access, capabilities, and outcomes (Wei et al. 2011; Scheerder et al. 2017) are unequal – characterizing the phenomenon of the digital divide. Lack of access to certain data can magnify inequalities (Shah et al. 2022), similarly lack of access to ICT can also escalate inequalities (Van Dijk 2005). The digital and divide are intrinsically connected, as in literature concerning the data divide digital access is highlighted as a key cost of data for people living in poverty (Mehrabi et al. 2021). Arguably, the data divide could be a part of the digital divide in that data analysis related capabilities are connected to digital capabilities, and outcomes of ICT usage are also outcomes where data is used.

Concluding Digital Divide

Thus, the digital divide in theory is quite complex and has many considerations that need to be made. Considerations range from the stakeholder for who the digital divide is being examined, different levels of the digital divide (individual, organizational, global), the different levels of the digital divide as determined by access, usage and outcome. While there are many factors that must be considered, there is not a single framework which clearly outlines where one level of a digital divide begins, and where one ends. This makes it challenging to scope and define divides through set frameworks as case by case there might

be different measurables to account for with the digital divide. However, it is important to acknowledge that digital divides exist and look at where they may be found in different contexts. In the context of this thesis, I will be examining digital divides in and between cooperatives in India. However, it will be important to examine the individual characteristics of the digital divide there as there are many parts of the digital divide under broad levels such as access, capability and outcome which may be essential towards understanding the case.

Outcomes of ICT and Data Use

The outcome divide was briefly explained in the digital divide chapter of this thesis. As mentioned, the outcome divide is difficult to define. Scheerder et al. (2017) found that although internet usage outcomes were not frequently studied the ones studied most in order were social, personal, economic, and motivational determinants (p. 1614). Thus, to further understand the implications of ICT usage, I will be studying outcomes of ICT usage and digitalization in this chapter. The popular view is that developing ICT capabilities is essential towards development of rural areas and the global south (Cullen, 2001). This section of the thesis will re-iterate why building of these capabilities is important, and how it can benefit people in their day-to-day lives. Additionally, this section will discuss some of the challenges that come with introducing technologies and risks that come with technological adaptation. It is important to consider outcomes as studying them can lead policymakers to make decisions towards egalitarian internet usage (Scheerder et al. 2017, p. 1614).

Technology, data, and development

The digital divide has been seen as a division between privilege and underprivilege, where those who are unable to or choose not to use ICT are seen as not being able to participate fully (Cullen, 2001). Data-driven technologies can improve decision making in farming, leading to better outcomes such as higher crop yields (Mehrabi et al., 2021 p.154). However, the implementation of these technologies is unequal due to differences in infrastructure (ibid). This highlights the potential of improving livelihoods and inducing development outcomes through technology and data.

Adoption gaps

Adoption gaps exist with many different types of technologies. These gaps can lead to asymmetrical outcomes. Carter et al. (2020, p.255) find that AI is beneficial to those who can

use and capitalize on the technology, and disadvantageous to those who do not have the skills to use it effectively or in general lack access to it. Carter et al. (2020) find that late adopters of AI may experience decline in competitiveness. As AI develops faster than policy and regulation (ibid.), this can perpetuate societal problems. Carter et al. (2020) further suggest additional research into AI related inequalities regarding access, ability to use, and outcomes of AI engagement.

Additionally, when new technologies are introduced, there can be skepticism about this new technology born from worries about the reliability of the new technology and the security of data with the new technology (Jakku et al. 2019). This can occur even when there are perceived benefits of big data and technology in general (Jakku et al. 2019).

Digital and Data Dispossession

A potential challenge in the usage of ICT is the potential for digital or data dispossession. "Accumulation by dispossession" iterates the concept that accumulating surplus profits through "dispossessing" individuals of these earnings (Harvey 2004). Zuboff's elucidation and coining of the term "digital dispossession" is the commodification of behavior, and the barring of access and rights to this now commodified behavior (Zuboff 2019, p. 100). The behaviors mentioned are ICT usage and data collected about and through these behaviors which can then be commodified by other entities to profit-seek. Thatcher et al. (2016) write that profit-seeking motivations are the driving force between data shifting towards an epistemological orientation, creating the danger that the quantity of data is taking priority despite what the quality may with a profit seeking motive. Gray writes that while scholars such as Zuboff (2019) make pertinent points in studying data commodification and extraction through "Accumulation by dispossession", they are neglecting to study inequalities in impact and input (Gray 2021). Gray views data dispossession as a way of reiterating "raced, gendered and classed modalities of power" (Gray 2021). Christian (2019) also cites examples about ways in which AI can be detrimental towards certain groups of people, due to the way data has been collected and processed. Proper data governance procedure is required to collect data that is accurate and representative, this will be discussed further in the data governance section.

Hassan argues that digital dispossession has occurred prior to accumulation through the creation of a virtual space (Hassan 2020). By this Hassan concurs that digital dispossession does not result in the same kind of feeling of loss as for example dispossession of land, meaning that the effects are not yet fully understood (Hassan 2020). What was previously private, is now commodified and privatized through asymmetrical power relations (Thatcher et al. 2016). As digital dispossession is an unknown disruption, it must be studied to understand the extent of dispossession, and the effect of being dispossessed from something that the loss of cannot be felt. Dispossession is recursive, meaning that it not only means that there is a transfer of property, but also that an object is being turned into property (Gray, 2021). From a digital standpoint, this means that data and the digital space becomes property and is transferred away from the hands of the owner/the creator of the data.

Concluding ICT Outcomes

Thus, to conclude this section it is important to understand the different outcomes of adopting technologies to further understand the digital divide and the consequences and benefits it can have. Understanding the consequences and benefits of ICT in cooperatives from a literature perspective from multiple levels (including individual, community and societal) grounds the purpose of this thesis and will assist in identifying benefits and risks for this thesis.

Data Governance

Governance is difficult to define, as many different governance systems with varying definitions exist (Ruhanen et al. 2010). Rhodes (1996, p.660), had a similar conclusion even remarking the term governance had "too many meanings to be useful". According to the definition Rhodes (1996, p.660) provides, "governance refers to self-organizing, interorganizational networks". This incorporates steering mechanisms also mentioned by Stoker (1998), while using a similar definition to the scope of where governance exists as Rhodes. Rhodes suggests that for governments the greatest challenge is encouraging the networks to discover new ways of co-operating. However, with the shifting of responsibilities mentioned by Stoker (1998, p.21) it is not clear what the role of the government may be in systems that act more independently of the government, and if the government should fortify their role in these situations. Stoker (1998, p.20) advises that to hold power in governance in the long run, this power must be seen as legitimate. This concurs that trust and legitimacy are crucial in building a working governance system. As with governance in general, data

governance is challenging to define as different approaches towards data governance exist (eg. data cooperatives, data trusts). Weber et al. (2009) highlight the need to understand that data governance is not universal, and organizations require individualized approaches, this underlines that there are many ways to structure and define data governance. For the purpose of this thesis, Janssen et al. (2020) definition will be used. Data governance is defined as follows:

"Organizations and their personnel defining, applying and monitoring the patterns of rules and authorities for directing the proper functioning of, and ensuring the accountability for, the entire life-cycle of data and algorithms within and across organizations." - Janssen et al. (2020, p.2)

This definition of data governance allows for examining data governance for the entire life-cycle of the data, as well as examining the outputs of the data such as algorithms used for AI. According to Micheli et al. (2020, p.6) designing data governance and deciding which elements to include in an organizations data governance approach is crucial towards effective data governance design. Data governance creates and assigns responsibilities, establishes standards across organizations, and ensures compliance with both strategy and local regulations (Weber et al. 2009, p. 2). Thus, data governance must be tailored towards organizational needs and decision-making styles to be effective, rather than using a data governance approach that already assumes certain roles and functions in the organization (Weber et al. 2009).

Data Production in India

Hanbal et al. (2020) conducted a case study about the process of data production in government information systems in rural India. According to them data protection is in the control of the local elite, meaning that it is critical to understand the dynamics of data protection and collection, as the way this data is handles provides an "illusion of transparency" (Hanbal et al. 2020, p.485). This is because there is a "political nature" (Hanbal et al. 2020, p.485), to producing data. It is crucial to understand the politicization of data production to be able to fairly evaluate the data and create better policy from it. Data production can have technological challenges as well and these should not be overlooked, but

the political way in which data is produced needs to be acknowledged as being reflective of the power dynamics and interests of activists from civil society organizations, local leaders, or officials (Hanbal et al. 2020, p.492).

Data cooperatives

Data cooperatives are a way in which members can band together their data to gain valuable insights about, for example, their current economic, health and social conditions (Hardjono & Pentland 2019, p.2). The fundamental aspects of a data ensure that "Participants of DCs share data while retaining control over it, having a say on how it is managed and put to value, and not submitting to the extractive logic of digital capitalism" (Micheli et al. 2020, p.7). The "extractive logic of digital capitalism" (Micheli et al. 2020, p.7) was discussed previously in the digital dispossession section of the thesis. The concept of data cooperatives stems from the cooperative movement to combat capitalism (Micheli et al. 2020, p.8). Data cooperatives differ from data governance structures such as data sharing pools, as they aim to create value, rather than capitalize on members to maximize profits (Hardjono & Pentland 2019, p.3; Micheli et al. 2020, pp.7-8). Data sharing allows people to become empowered and solve local problems, through overcoming the concentration of power in the data landscape (Wu et al., 2021). This promotes the idea of data-sharing in cooperatives as a way for cooperatives to become engaged in the data landscape, which is a way for cooperatives to combat the privatization of behavioural and other data.

Data cooperatives are based on fairness and trust of stakeholders involved (Ho & Chuang, 2019, p. 203; Micheli et al. 2020, pp. 7-8). Fairness is based on the distribution of the surplus between those who have contributed to the creation of value from data (Ho & Chuang 2019, p. 203). Fairness in distribution of this surplus is a fundamental appeal in participating in a data cooperative, as in this way data can be given at will. However, even in this situation where data is gathered consensually, viewing data collection as "benevolent" is dangerous as consequences of data collection can be widespread (Crawford 2021, p. 121). Trust in data cooperatives is necessary as members are trusting the governing mechanisms of the cooperative to function in the agreed upon terms. Members must trust that the way in which data is being collected, processed, and managed is fair and equitable to all. In return, the data cooperative must function in a way that reciprocates this trust and avoid misuse of data.

Data cooperatives create value for their members through data, even in areas where the communities do not understand the benefit (Scholz & Calzada 2021). The data cooperative model allows for protection of community data rights and ensures data fairness and a focus on social benefit for members through the usage of this data (Scholz & Calzada 2021) The value data cooperatives create can be subtle, and sometimes are not seen by the affected communities (Scholz & Calzada 2021).

Data Trusts

Mulgan & Straub (2019) argue that an ecosystem of trust is needed to ensure that data is both trustworthy and trusted. Mulgan & Straub (2019) explain that we should expect different forms of data trusts to form in the future, the form of these trusts could range from sharing data, to managing data, or even providing research capabilities. However, the challenge is to create trust in these data trusts (Mulgan & Straub 2019) in order to encourage member engagement, as well as build accountability for these trusts either to the members of the trust and in some cases to the public as well (ibid.) Data trusts differ from data cooperatives in that decisions made with data are made by a board of trustee's with fiduciary responsibilities, while data cooperatives have member driven democratic decision making (Bühler et al. 2023, p.150).

Wu et al. (2021) write about how data governance technologies can be used to democratize data sharing for community well-being. Data concentration is shifting from government to public sector (Wu et al. 2021). This shift in the collectors of data, means differences in both the nature of data, the frequency of data collection and the usage of this data.

According to Wu et al. (2021) lack of data literacy and lack of data skills are a major barrier towards data sharing in data trusts. However, when referring to these skills Wu et al. (2021) assume a high level as skill, as the data literacy mention is complications in operating data trusts, not merely contributing to a data trust. Thus, for data trusts and other data sharing functions to operate data literacy is important at all levels of data governance, from participation to those operational management.

Concluding data governance

From reviewing literature about different modes of data governance, various challenges towards data governance and different factors affecting this data governance: some conclusions can be made. Firstly, there are data governance models beyond the data cooperative model in which cooperatives could store and utilize data. The data trust model where a body of experts with fiduciary duties control the data could potentially be used. Secondly, through this section, and the ICT outcome chapter the importance of collecting data that is unbiased, or being able to analyze data critically is underscored. The collection of clean data trickles down the value chain of different kinds of data governance and affects decision-making. Thus, data governance is not universal and should be considered case-by-case what the best approach to achieve reliable and usable data is.

Research Gap

To conclude the literature review, I will now present the research gap that emerges to further ground the purpose of the thesis. Data governance, cooperatives, and the digital divide are all topics which are gaining traction and being covered by literature. However, a gap exists in examining the trade-offs of data collection for cooperatives in a setting where there is also low digital literacy. While Wei et al (2011), present a framework which examines the digital divide as a set of access, capability, and the outcomes of the divide.

Meanwhile, Van Dijk (2005) presents the digital divide in five steps as presented previously. These frameworks acknowledge that the digital divide is perpetuated by a combination of the factors identified in the frameworks interacting, in the end create some sort of output. In Van Dijk's (2005, p.15) "causal model of resources and appropriation theory" (Figure 4) access to ICT is shown to affect the impact of participation in society, and thus further inequalities. Wei et al. (2011) focus on skills and knowledge outcomes in their study. However, in outcomes of ICT and Data Use chapter suggests that digital dispossession (Zuboff 2019) and adoption gaps (Carter et al. 2020) can be taken as risky outcomes when discussing data and digitalization, while there is also potential for improvement in things such as decision-making in farming (Mehrabi et al. 2021).

Data cooperatives and data trusts rely on trust and participation of members in sharing data (Ho and Chuang 2019; Micheli et al. 2020). However, member participation in cooperatives can be affected by many factors such as the size of the cooperative, and varying interests inside the cooperative (Österberg & Nilsson 2009). Cooperatives can also have different levels of market-orientation which can also change over time (Bijman & Wijers 2019). Existing literature about data cooperatives seems to mainly focuses on cases where data and digital services are the primary function of the cooperative (eg. Bühler et al. 2023), however the transition of existing cooperatives into this sphere is an interesting phenomenon.

Realization of outcomes and data governance decisions in cooperatives are affected by the digital divide. The digital divide is a phenomenon that is studied in the context of many countries and some demographics of people are affected more than others (Singh, 2010; Rajam et al. 2021; Lythreatis et al. 2021), however, there is a gap of current literature exploring how cooperatives data and digitalization efforts are affected when there is a

differing or low ICT knowledge between members. Although scholars such as Bühler et al. (2023), do mention digital exclusion as a challenge to be addressed by data cooperatives trade-offs of data and ICT usage in cooperative settings should be further studied. The realm of data cooperatives, and data trusts are only just gaining traction in literature. Thus, the function of existing cooperatives in harnessing data and where the risks lie in this should be explored further.

Thus, the research gap is the uncertainty of the outcome divide of ICT and data usage, and the ability of people with low digital and data literacy to benefit from digitalization and data sharing through cooperatives – and what additional risks they are exposed to with these developments.

3 Methodology

In this section of the thesis, the aim is to explain the methods engaged to gather data. Furthermore, the aim of this section is also to introduce the methodology of how the data collected has been analyzed. Additionally, I aim to present the rationale behind choosing the selected methodologies. Finally, as with all forms of research, there are some limitations to the methodology selected as well as ethical considerations which I have presented.

Choice of Research methodology

I chose to conduct my research using a qualitative approach. A qualitative approach is useful when exploring a topic. Patton (2015) mentioned several reasons why one would choose to conduct qualitative research, and from these reasonings I chose to conduct qualitative research as I am aiming to explain the function of a system and how it affects people (Patton 2015, p.13) and understand how and why context matters (Patton 2015, p.13). These goals are highlighted in my research questions, which were:

RQ1: What are the trade-offs of collecting data in cooperatives?

RQ2: What are the digital divides faced by cooperatives?

RQ3: How is data governance in cooperatives affected by the digital divide?

Here I am researching to both contextualize and explain a phenomenon, and how this phenomenon impacts people. I am examining both how data governance and digital divides arrange themselves in cooperatives, as well as how this effects cooperative members and understanding the context of a specific case to gather information. Thus, because the nature of my inquiry is open-ended and seeks the kind of results mentioned before, I have chosen a qualitative study. Additionally, since I am not measuring views but rather investigating them, a qualitative research approach is more appropriate than a quantitative research approach for my thesis. My thesis takes a pragmatic approach, as a pragmatic approach is described as one that is problem centric and aims to investigate the what and how (Creswell, 2009).

There are several types of qualitative data. The qualitative data my study is derived on takes form of interviews I have conducted, interviews my peers have conducted and a baseline report conducted by Self-Employed Women's Association (SEWA), a cooperative federation.

Case study

The way in which case studies are bounded is arbitrary, and there are different schools of thoughts about whether a case study is a "bounded entity" or a "bounded system" (Patton 2015, pp. 259-260). Patton even writes "The variety of approaches to defining a case gives you an opportunity (and responsibility) to define what a. case is within the context of your own field and focus of inquiry" (Patton, 2015 p. 260). Thus, I will be following Creswell's (2009) approach of the "bounded system" as a case study. I will be studying the data collection and usage in cooperatives in India as a case, through the perspective of experts in the area. By bounding this as the case, I was able to sample purposefully to collect data which is relevant towards my research topic.

Data

In total I conducted three interviews. These three interviews were conducted with experts from cooperative federations, which represent and aid cooperatives. A total of six individuals were interviewed. Additionally, I was able to access two interview transcripts of interviews with members of a cooperative. In the secondary interviews, eight cooperative members were interviewed. I was also provided access to a baseline report from Self-Employed Women's Association (SEWA), which provided further information for the case.

Interview

The type of interview I decided to go for was a semi-structured interview. A semi-structured interview is one in which there is a guide of questions or themes for each interview, but no strict guide with guidelines of what questions to ask. I wanted the interviews to be semi-structured in nature as I am interviewing about a topic which is quite new to me, and the people I was interviewing were experts. This meant that all the people I interviewed were more skilled and advanced in the topic which I wanted to find out about, and thus had a wealth of knowledge they could share. Because of the expertise I decided to conduct semi-structured interviews to be able to flexibly learn more about the experiences and ideas of the experts in case they brought up topics I had not prepared questions about or emphasized the importance of some interview questions over others.

Sourcing Interviews

The initial interviewees were sourced interviewees through my thesis supervisor. An initial meeting was arranged informal introduction to a cooperative federation to learn more about the organization. After I had created an idea of the topic I wanted to research, I was able to arrange a brief meeting with the research coordinator at the cooperative federation. Subsequently, the research coordinator was able to support me in arranging an interview with experts familiar with the topic at the federation. After conducting this interview, I asked interviewees if they had any ideas of other people that I could speak to about the topic or similar topics. From this I was able to get the names of some further people to contact. This technique is called "snowballing" (Patton 2015). The purpose is to find more people to interview through a network. I then contacted the other experts I had been referred to and was able to arrange subsequent interviews through snowballing.

Interview Conduct

The interview guide I created consisted of 5 over-arching questions, with some sub-questions. The purpose of the guide was to help me as an interviewer during the interview to provide some general probes about topics I was interested in hearing the interviewee talk about relating to the research objective. The first version of the interview guide can be found in Appendix 1, however it should be noted that as I interviewed questions changed although they followed similar topics to the interview guide. The intent of the interview guide was not to serve as a list of questions. As I conducted more interviews, I noticed which questions worked and which ones did not and adapted as such.

I made a point to end each interview with an open-ended question. From both my past experiences in interviewing for qualitative data collection and asking a question such as "is there anything that I didn't ask that I should have or any comments you would like to make about the topic". The concept of having such an open-ended ending questions is also recommended by Patton (2015, p.470).as it can provide great additional data. This practice allows interviewees an open-form opportunity to add themes and topics that they might find relevant towards data governance, cooperatives, and digital divides that my questions didn't highlight during the interview.

All interviews were conducted and recorded via Zoom.

Analysis

To analyze my transcribed interviews, I decided to code the interviews thematically. I came up with the codes both inductively and deductively. The codes that came inductively meant that the codes were generated through reading the interviews and deriving different thematic patterns which emerged from what different interviewees said. Naturally, from my literature review and the very topic that I set out to explore some of my codes were deductively formed. This meant that the codes emerged from the defined topics and I created them for the purpose of answering my research questions and being able to group data together more clearly.

Figure 5: Circularity of Coding (From Decuir-Gunby et al., 2011 p.138)
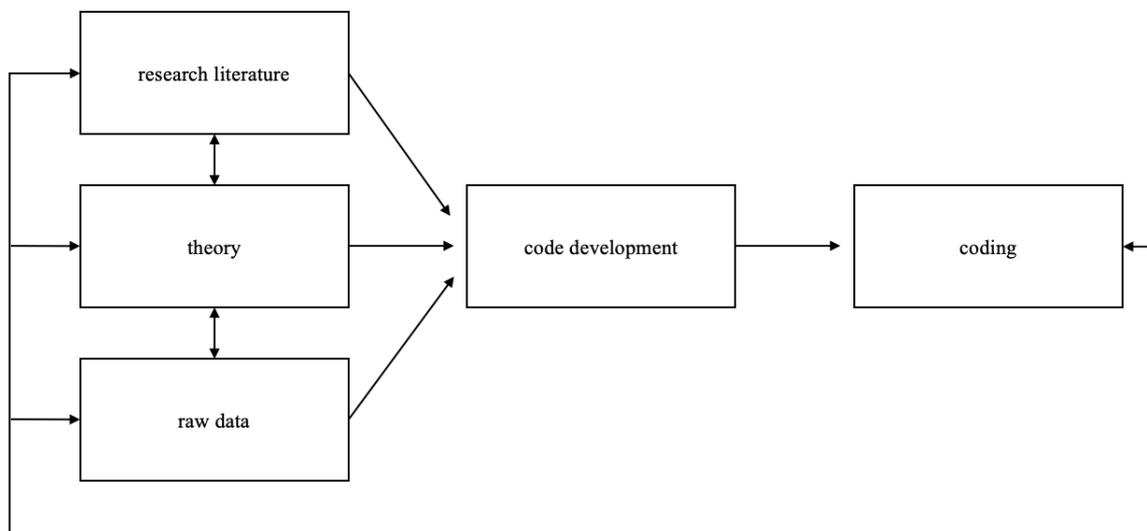
From the themes identified in the interviews, and the codes I created I then proceeded to make a codebook. The purpose of the codebook was to track how I planned to code the interviews and to clarify even for myself what each code meant. Codes are descriptions used to help analyze data (Decuir-Gunby et al. 2011, p.138). According to Decuir-Gunby et al. (2011), often times the coding process is not described in research and not following a clear systemic structure. To code my interviews, I followed the structure presented by Decuir-Gunby et al. (2011), which included first identifying codes through literature and from the interviews. Then from here on the codes identified from the interviewed were grouped as large overarching themes, thinking of these codes as large themes helped me break down codes that were too large into more manageable codes that were at this point categorized into larger overarching themes. I then organized these themes and the codes underneath them into

a table, which became my codebook. The codebook can be found in Appendix 1. The process in Figure 5 suggested by Decuir et al. (2011) was followed, making the coding an iterative process.

For example, a code such as "trust" could potentially mean multiple things such as the interviewee having trust that something is happening or the interviewee speaking about trust being something that emerges or is needed in data governance. When coding, it is important to be able to make this distinction to code more accurately. The codebook I created can be found before, it includes each theme, whether it is inductive or deductive, a description of the code and a quote that is an example of the code from an interview (this was removed from the thesis for redundancy purposes).

I conducted coding with the use of Atlas.ti, a qualitative analysis program, where I uploaded all interview transcripts and coded on the basis of my codebook. This allowed for me to sort, organize and find coded text later when writing up my findings and discussion. Since I was the only one coding the interviews, I did not need to check for the consistency of my coding or train anyone else to use my codes.

Limitations of Research Methodology

As with all research works there are some limitations and potential biases that occurred in my research methodology.

One of the main limitations often cited for case studies is their lack of generalizability. Case studies focus on studying a specific phenomenon at a specific place and time, and thus can lead to a niche understanding of the topic from one perspective. My sample however does include interviews with experts outside of the case organization, leading to a broader view of the topic.

Another limitation is that the interviews were not conducted in the interviewers or interviewees common native language. According to Patton (2015, p.481) there can be misunderstandings related to language. However, from my perspective this did not turn out to be a major challenge as we were able to communicate clearly, repeating and rephrasing when necessary.

The interviewers' skills are important in gathering the results (Patton, 2015). While I cannot myself conclusively evaluate my own skill as an interviewer, I will say that it could potentially be limited due to my lack of prior engagement in the topic and experience interviewing. However, I was able to develop findings with the data gathered.

Ethical Considerations

When conducting research there are always ethical considerations to be made. I ensured that my interviewee's were familiar with the topic they would be interviewed on and had the opportunity to ask questions regarding my thesis to establish informed consent, I ensured that interviewees knew their rights to privacy, and finally I also ensured that interviewees would remain anonymous.

Informed consent is important when conducting research. To conduct the interviews ethically, prior to the interview I sent out a research privacy notice summarizing the type of data that would be collected from the participants as well as. A consent to participate form, to confirm that the participant was aware the interviews would be recorded and used for this master's thesis project. In addition to this, prior to every interview I asked for verbal consent once again to record the interview, to ensure that the interviewees were okay with this.

In obtaining informed consent, when reaching out to potential interviewees unfamiliar with the thesis project I gave a few sentences summary of the topics I would like to interview them about and wrote that I would be willing to answer any questions they may have on that topic. At the beginning of the interview, I also gave a summary of what I was interested in, and at the end of the interview I also asked if there were any questions. Additionally, I sent the interview guide to participants only upon request.

I created random identifiers for each person interviewed, such as Interview Participant X (IPX) to preserve the anonymity of my interviewees. Ultimately, in writing the thesis I decided not to disclose the participants positions within their organizations or the organizations they belonged to, as I felt that it would compromise the anonymity of the participants.

4 Findings

In this section of the thesis, I will present my findings from the interviews I have conducted. I have categorized these findings based on the codes identified. Each topic will be presented with general findings and example excerpts. I will classify interviews conducted by my peers as "secondary" interviews to distinguish them from the data I collected myself. Additionally, the SEWA baseline report is a document provided by the SEWA federation.

Background of findings

In order to contextualize, I will be briefly explaining the kind of data projects discussed by each interviewee and the baseline report.

Interviewees IP1-IP3 are primarily discussing a project involving collecting data from women farmers to create pricing algorithms. The SEWA baseline report is a study on the current technological situation at the regions the organization is involved in. Additionally, the two secondary interviews with Secondary IP1-IP8 are conducted with a producer cooperative.

Interviewee IP4 is mainly discussing the Indian Governments cooperative digitization of Primary Agricultural Cooperatives initiative.

Interviewee's IP5 and IP6 are discussing a multitude of data-based initiatives for cooperatives.

Introduction to Findings

Collecting data and adapting ICTs can have many opportunities and benefits for cooperatives, however, challenges to the process also exist. When cooperatives digitalize and collect data to benefit their members, new risks and opportunities arise. Thus, trade-offs exist when the decision to collect cooperative data is taken. Additionally, even before this there are existing challenges in being able to access some specific cooperative member bases to collect data.

"We know that data when it is digitized is most useful." – IP1

There is a connection between data and digitalization, as in a digitized form data can enable more effective outputs, however as mentioned there are trade-offs also involved in the steps that come along with this which is the collection and the usage itself. Thus, the findings are structured through key areas of the data collection process identified through both the findings and inspired by the literature review – followed by the of risks and benefits expected from data collection and digitalization in cooperatives. Firstly, the findings will go through key areas that arose when data is being collected. Secondly, the findings will go through themes that arose when the data is being made usable, ie. Data analysis. Finally, themes that arose through outcomes of access to data and ICT are presented. The process of the findings section can be seen in Figure 6. The digital divide is present in every stage, thus it is depicted at the bottom of the diagram.

Figure 6: Structure of Findings
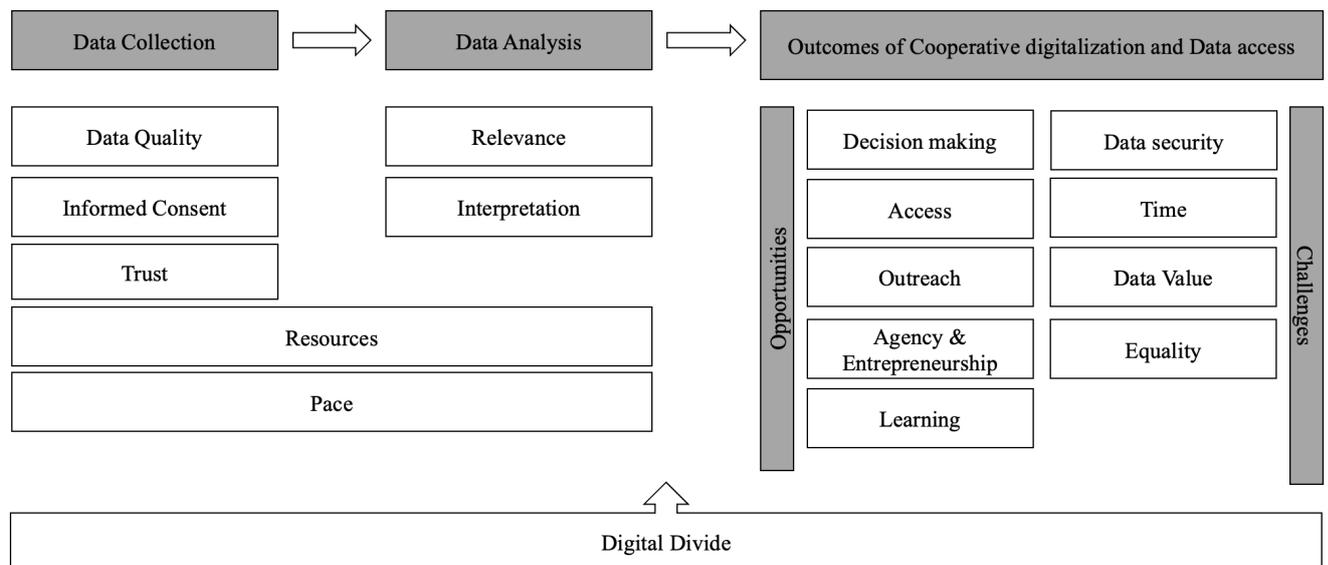
Data Collection

When thinking about data collection in terms of cooperatives, specifically in the India context, several themes and ideas emerged through the interviews. Firstly, relating to data collection the question of who is the one collecting the data emerged. Secondly, how data is being collected is also important, as different approaches may be employed. Thirdly, what data is being collected and its relevance to those who it is being collected from arose.

Data Quality

When collecting data, there is a big question of "who" is the one collecting data, as there is a danger of biases pouring in. As mentioned before, biases in data can lead to a multitude of consequences when the data is used (Janssen et al. 2020). IP4 acknowledges that biases may even impact who is selected to be collecting the data.

"in India especially cooperatives being a state subject and especially agriculture has a lot of biases in place on religion, on caste, on gender, and various sorts. So these kind of biases do pour it in and there's sometimes the participation is very much mobilized by such biases within the local communities to see who will be an appropriate person collecting or monitoring or gathering this kind of data." – IP4

Cooperatives are crucial to collecting data from populations with low access to ICT. This is because cooperatives are recognized as some of the only local institutions that are trusted by their members and communities and have the presence in even some of the more remote places in India.

"These {cooperatives} are the only entities which have good presence and good governance and good understanding of the local environment in their local communities" – IP4

Some interviewees spoke about the hybrid and solidarity models in cooperative data collection. The hybrid model as explained by IP1 means that not everything in cooperatives is digital, but rather that information is also provided through trusted sources. These trusted sources are grassroots leaders, board members, public leaders and leaders in the cooperative federation; connecting members who otherwise would not have been connected to the information because of the solidarity model (IP1). Because of the unity – solidarity –

members are not left behind due to low digital literacy levels, as the responsibility to share the information is communal. This creates an opportunity where cooperatives can bridge even those members with lower digital literacy to create value through the data shared, and data can be collected through trusted members of cooperatives. However, with the problems IP4 mentioned, there is also a risk that data perpetuates societal challenges, the data may be potentially harmful for users as well.

"What we follow is a hybrid model so everything isn't digital" – IP1

Due to challenges in areas such as network connection, the way in which data is collected is important. This is because data collection is not straightforward and requires specific tools to be able to do it. Some of the tools being used to conduct research was KOBO tool for grassroots research (IP2), however there are still challenges with literacy to collect data.

"in a few of the cooperatives the situation is that they understand that the data is important, but how to collect the data how to do this, is still a problem for literacy data collection is not up to that level." – IP3

Informed Consent

Informed consent is consenting to something that one understands and knows the meaning and repercussions of. Throughout the interviews, the idea of informed consent was mentioned by participants explicitly and vaguely.

"consent is, is something which is really missing and to be able to give consent you need to know what it is, you know, what you're consenting to and what consent actually means. So, I don't think that is the case in the majority" – IP6

IP6 brings up concerns about the lack of informed consent when data is being collected. Due to low digital literacy levels, the understanding of what is happening with data can also be low leading to concerns about informed consent. Lack of informed consent is also an issue from an organizational perspective, as there is a concern that consent in data usage was missing due to a lack of understanding of both the value of the data, and what would be done with the data. IP6, highlighted that this is a general issue in India, and not limited to

cooperative members. However, cooperatives are in a position where they can educate members about usages of their data and the value of the data.

Trust

From the cooperative member side, through the secondary interviews it was evident that there seemed to be an explicit trust in the cooperative and cooperative federation to share data/information. However, this level of trust implies there may be a lack of informed consent – as members are willing to provide the cooperative and federation with their data with full trust without questioning the intentions of how their data would be used when provided to key sources. The question remains, if digital and data literacy is low how can informed consent be given for data usage – and whether this is overridden by the benefits of the usage of the data. Trust is a double-edged sword when it comes to data collection in cooperatives. On one hand, trust allows for cooperative data governance mechanisms to work as there needs to be trust in allowing data to be used and in the data being managed by cooperatives and other actors collecting it. However, on the other hand when there is trust links very much to informed consent and how care should be taken when thinking of who to provide data to.

"All the data used for the benefits of the members only and they have a high trust in sharing with the cooperatives so they do share" – IP2

"we trust them and share all our information with them as they work for us. There are not trust issues here. But, if an unknown person approaches us for information, we do not share information with them" – Secondary Interview IP4

Secondary Interview IP8 explains that when information was asked for through the cooperative channel, there was complete trust in providing this information without concern for how the information would be used. This reiterates, that trust is key in the consent for sharing data and information. However, the trust is not extended to everyone and is mainly for government entities such as KVK's, the cooperatives and cooperative federations – organizations which cooperative members all see as working for their benefit.

Data Analysis

For data to be useful, it needs to be analyzed so that conclusions can be drawn from it. However, in contexts where digital literacy is low there are concerns with this.

Relevance

There is an importance towards what kind of data is chosen to be collected. The selection of data depends on the purpose the data is intending to fulfil. The quality of the data is affected by the feasibility of the data collection, and the relevance of the data is important for analysis for the data to be usable.

This implies that even with data collecting practices wherein which may in place, there may be challenges with data quality in the process.

"So yes they do collect the data, but in, you know, not in a very systemized manner. Which is where the major challenge lies right now and - how do you monitor this data? or make sense of this data, analyze this data and use it for future policy making - is what the country or the cooperatives in the country are working towards right now.". – IP4

Different projects have different reasons and purposes for data to be collected. Additionally, due to scale of some projects, it can be difficult to collect the data in a systemized way.

"we mostly collect data when it's needed and help in taking relevant decisions"– IP2

This indicated that the interviewee was not concerned that data that was unnecessary was being collected. The interviewee expanded that:

"because we already have some difficulties in collecting data, we only collects stuff that we know is going to be useful, not just you know collect it and then try to find a use for it" – IP2

However, despite this being the intent another interviewee states that:

"So, my one point is use of data when we talk about the content, about the data selected." – IP3

IP3 indicates that although only relevant data is collected, there may still be concerns as to whether the costs of the data collection outweigh the benefits of collecting the data. This is because the data collected is collected with a purpose to be used by cooperatives and their members. However, although the aim is to choose to collect only relevant data, it can be difficult to pinpoint exactly what the relevant data is, and to effectively explain the relevance to members to explain how data collection is useful to members.

"top-down approach here. It went through a lot of consultation and workshops at the national level and at the state level to see what data is important and what data is historically present, what data can be generated, and it's viable to be generated if not already, and of course what data will be important." – IP4 about the digitization of Primary Agricultural Cooperatives

IP4 describes a process done to arrive at what kind of data is selected to be collected. In this case the way of selecting the data is described as "top-down" however, IP4 also mentions a point about the feasibility of generating the data – which is important to realize given constraints such as resources which will be discussed in a later section.

Interpretation

"Once the data is there the main concern is usability." – IP3

"Also, they don't know how to analysis the data and what the data says all about" – IP2

Once data has been collected, the data should be able to be used to take decisions or to become better informed. However, this is challenging as once data is available it requires analysis. Without knowledge of how to analyze and interpret data, the data is not usable.

Data Collection and Data Analysis

As was portrayed in Figure 6, pace and resources were themes which came up both in relation to data collection and data analysis.

Pace

Data collection can be slowed due to difficulties such as poor network connection. In general, a slower pace in data collection can affect the quality of the data as it is then older when it is used – meaning that information may be outdated.

"Because of the- the lack of access that women have to digital technologies that is there is say a challenge in the pace we can't collect data as fast as we can and once we do collect it, maybe it's become redundant, right. Because for businesses you need accurate data really fast so you can make a decision based on what the data is telling you and- and therefore you know sometimes it- often actually- we miss out on changing with the market or women aren't able to, you know, pivot quickly because they just don't have the information" – IP1

Members' lack of access to technology also creates challenges for data collection due to first-level digital divide barriers (as defined by Wei et al. (2011) and Riggins and Dewan (2005)). Due to the lack of access to digital technologies that some of the members of cooperatives have, data collection can be difficult as the lack of access impacts how quickly data can be collected. IP1 mentions that pace is crucial to have up-to date data as when it is used for business decisions, it becomes redundant when not up to date.

Pace in data collection may also be hindered due to barriers in network access.

"many times we lose the connection if we are collecting the data using the digital technology" –IP2

Collecting data digitally is only possible when there is some level of access to digital technology by those whose data is being collected. Thus, the medium in which data is collected is important as it determines from whom and where the data can be collected.

Resources

Resources are needed to both be able to collect data and analyze the data to achieve desirable results.

"human resource to maintain and manage the data" – IP2

Resources for being able to analyze the data are scarce, especially due to lack of digital and data literacy. Additionally, resources in managing data are also limited. However, this is where the cooperative federation can support with different kinds of trainings if members find it relevant.

"They [cooperative federation] asked us members what our needs were, and that they wanted to train us on google meet, zoom whatsapp, and they asked us if they wanted to be trained on this and if it was important to them." - Secondary Interview IP4

Outcomes of digitalization and data usage: Opportunities

There are many different opportunities that arise due to digitalization and data usage, in this section of the findings some of the opportunities identified will be explained.

Decision making

Improved decision making is a benefit of data collection by cooperatives. Through the collection of data, the aim is to improve the ability of decision-making through using data.

Participants highlighted the specific importance of data from cooperatives being used by the cooperative members themselves as a kind of check for cooperatives and reinforcing the cooperative principles. Cooperatives already feature democratic decision making, thus with data these decisions could be more informed for members. Additionally, through access to data members can ensure that decisions are being made that keeps cooperatives as cooperatives. Thus, a perceived outcome that is expected through using data in cooperatives is that better decisions will be made also in the context of the cooperative and its future direction.

"democratic decision making is what cooperatives are well you know for. I mean there's seven principles of cooperatives of autonomy and independence of control by the members decision making by the members… I think they [cooperatives] *will benefit a lot from having more accurate effective data to make these decisions, which of course in the long run can help in providing the right kind of environment to the members of these cooperatives: socially, environmentally, economically."* - IP4

"making sure that cooperatives do not also cross a line from cooperation to corporatization that's also something the data can better lead and this better sort of awareness if you give to the cooperative members, I think better informed decisions will be taken" – IP1

Beyond decision-making using the own data collected, decision-making is improved when there is greater access to information which can come about from the usage of digital technologies. This is further discussed in the next section: Access.

Access

Access to both information and digital technologies can be beneficial for cooperative members as it can provide know-how and information (like weather forecasts) that improve decision-making.

A key benefit identified through both the secondary farmer interviews was access to weather forecasts. Farmers mentioned how beneficial it was to have access as weather forecasts are essential towards farming, as they can help to make decisions and be prepared for different weather conditions.

"We also use YouTube for more information and news and this has helped us a lot in avoiding weather related problems."- Secondary Interview IP5

"getting information ranging from weather reports and their to- to your, you know, credit balance." – IP5

The ability to access and use technology is foreseen to create new opportunities for members of producer cooperatives. These can be seen through both interviews with experts working with cooperatives, and the secondary farmer interviews. From a secondary interview: When a

farmer was unable to attend a training exercise, they were able to access the training online at a later time. This kind of access is crucial as it opens up more opportunities to receive information.

"Then, I also watch and upload stories on Facebook regarding certain farming activities that I undertake, or new methods I learn, and even share other people's methods." - Secondary Interview IP3

"We also have been trained on zoom, google meet, so we use these platforms for communication and video sharing." - Secondary Interview IP3

However, the adoption of new technologies for some cooperatives came out of necessity during the pandemic, during which access and ability to participate in cooperative activities was hindered as these events were traditionally in person.

"I think the last three years what we've generally seen is also because of the pandemic, not because of choice, but mostly because of circumstance. ... I think starting from online meetings, filing of-of you know notes, annual general body meetings, and you know all statutory requirements" – IP5, when asked about cooperative digitalization projects.

Thus, it is an interesting circumstance in which accessibility of meetings has partially moved to a virtual space also within cooperatives, where this was not the case prior to the pandemic.

Outreach

Although data collection is difficult to manage and organize in rural cooperatives, the flipside of this is that when data is collected by cooperatives there is a unique opportunity to obtain this data as cooperatives can provide a link through which to collect data. Data collection through cooperatives can reach a population that otherwise may not have been reached in a data collection exercise. Indicating, that despite difficulty in reaching cooperatives members – cooperatives are still in a better place to obtain member data than other entities The kind of data cooperatives have is unique towards the makeup of their membership and data which is present can be data such as gender ratio's, literacy levels, or even youth leader presence (IP4). All of this data can be helpful for policymakers to make informed decisions and monitor situations in the more rural outreaches.

"It's all present with these cooperatives because of the kind of membership that they represent." – IP4

" their {cooperatives} presence at, you know, the remotest parts of the country in India give that kind of uh enabling environment or-or a framework of actually monitoring all kinds of welfare schemes or mechanisms by which the government can reach these last-mile entities last-mile villages last-mile pockets in the country." – IP4

Data collection by cooperatives may allow for members to become better connected with services and opportunities they otherwise would not have been connected to – particularly from the government when speaking of the efforts to digitalize Primary Agricultural Cooperatives. Mapping and gathering data on cooperatives can also help better inform decision makers on policy, and the current situation in the more rural parts of India. Visibility can also be gained for cooperatives and their members through collection of data.

Cooperatives are important in "last-mile" (IP4) outreach as their members can be very rurally located and difficult to reach otherwise. These cooperative members benefit by being reached for data collection through opportunities to participate in government schemes, and their data being used to help make better policy decisions at a state-level. These schemes can include cooking schemes, solar and irrigation systems. IP5 mentions that in the case of migrant workers in India specifically, if financial and spending pattern related data can be collected, then there can be more financial services offered.

"...hence they [cooperatives] *are best suited in the in the view of the present government to be able to receive and share the benefits of these various government… hence the members do benefit a lot by being part of these cooperatives."* – IP4

"If these can be collected, a lot of financial services can be given to people at the bottom of the pyramid." – IP5

The improvement of services and creation of relevant policy can be enhanced through collecting data from cooperative members.

Agency and Entrepreneurship

When interviewing those involved or those observing implementation of data governance and management, it is important to remember the agency of cooperative members and their own right to decision making. The purpose of cooperatives is to improve social welfare of members and communities through providing a platform for democratic decision making. Thus, it is important to remember the agency of cooperative members involved in digitalization and data gathering projects. Through the interviews, it was evident that having access and ability to use technology was seen as promoting agency and creating more choices. Another way of looking at agency is seeing how an entrepreneurial mindset towards technological adaptation can carry over to change the way technology is used in cooperatives. Thus, this section of findings highlights the agency digitalization brings to cooperative members and how it can in turn engage entrepreneurial qualities. Through the data, the theme of entrepreneurship arose. Not only does access and ability to use technology provide cooperative members with low levels with more degrees of agency and self-sufficiency, access, and ability to use technology promotes entrepreneurship and provides the end-user with new ways to both develop and improve existing produce, as well as market it with further outreach.

"used youtube to watch videos on composting, making organic fertilizers and other information on organic farming." – Secondary Interview IP3

"These women who otherwise would not have connected to information and now connected to information because they're part of the solidarity model." - IP1

"And I mean finally ownership, because like I said owning the data is- and being able to make decisions about and using your data is crucial. So, these women are again being part of the solidarity model do have ways and means to do this now which they otherwise wouldn't" – IP1

This demonstrates agency of farmer themselves. With access and ability to use technology, farmers are able to access education on the internet and circumvent physical barriers that may have otherwise meant missing out on educational opportunities due to constraints with space and time. The ability of cooperative members to benefit from trainings online, creates new

opportunities that otherwise would not have existed. Additionally, because the women in the cooperatives own the data they are producing – this is a way in which data governing mechanisms in cooperatives differ. Ownership of the data allows for women to exercise more agency over decision making.

Entrepreneurship is typically characterized as taking initiative in a new business type. In this case, the entrepreneurship code was used to define activities taken by cooperatives, and cooperative members which demonstrated agency as well as activities that would typically be characterized as business activities. The usage of social media to promote activities was identified through a secondary interview as an entrepreneurial activity that is being undertaken, as members created an Instagram page for their cooperative.

"We use Whatsapp the most, and also Facebook for uploading stories on what we do. We are also starting an Instagram page for [cooperative]*."* – Secondary Interview IP4

However, the benefits of being connected to larger markets through access to ICT have yet to be fully realized as:

"94% of all respondents have never used online platforms for selling their produce." – SEWA baseline report

This is an indicator that outcomes of using ICT may not yet be at a larger level of benefit. However, the fact that a small portion of members had in fact sold their produce through an online platform demonstrates entrepreneurship as well as the potential of digital activation.

With digitalization comes greater opportunities and access to information and services. However, as of yet, the situation described the outcomes of digitalization are described as not being at the potential level they possibly could be. Digital tools are not yet completely harnessed for "productive work", however digital tools are in use for communication purposes.

"they [cooperative members] seem to use a lot of WhatsApp again for communication purposes but using a digital tools for productive or quote unquote productive work such as, you know, marketing and selling their products or services, getting access to information

about their businesses or their work is something which is still look at. So, at every level you see that women have been sort of kept away from realizing the potential of the digital economy of the internet and power that it brings. " – IP1

Learning

One of the benefits of both digitalization by cooperatives is the opportunity of learning. IP2 mentions that those with low digital literacy levels face many issues, thus cooperative digitalization may be able to bridge the digital capability divide. Having a low digital literacy level can mean lack of access to information, including about government schemes which may be helpful, lack of access to credit, as well as access to general information about weather and farming. The barriers to information are being circumvented by cooperative federations, through digital literacy programs and new platforms for sharing information. Additionally, in recent years there has been increasing digitalization in cooperatives – such as online meetings for example.

"we have found that the low digital literacy level faces many issues" – IP2

"I think the last three years what we've generally seen is also because of the pandemic, not because of choice, but mostly because of circumstance." – IP5, about digitalization in cooperatives

From the perspective of the secondary interviews which were conducted with cooperative members, training by the cooperative federation helped in members learning how to use a phone, and use it for activities such as accessing meetings, information and communicating with others. Additionally, learnings can be shared among the cooperative because the on solidarity model of cooperatives enables this and there is trust in place to do so.

"When the need of using smartphone arose after everyone started using it, I bought one too but did not know how to use it…then they [cooperative federation] *started training the Leaders on how to use the phone for accessing information – and then the leaders in turn trained the other women farmers."* - Secondary Farmer interview IP1

"... Now that I have training from [the cooperative*] on how to use the phone, message, use apps and join meetings – I have a phone myself and I use it a lot.*"- Secondary interview IP

Outcomes of digitalization and data usage: Challenges

Although many opportunities were identified through digitalization and data usage, there are also some themes of challenges that were identified in the findings. These challenges will be explored in this section of the thesis.

Data Security

The importance and the state of data security was brought up in interviews. The interviewees mentioned some major data leaks, and some of the issues surrounding them.   It was mentioned that in the context of India, data leaks such as that of confidential vaccine data had occurred (IP6).

"security or how strong is the data governance frameworks, how many leaks have already happened, these are all questionable. In the last couple of years many public deliberations have happened, many leaks that you know media has reported and even the highest actually owns of bureaucracy their own personal data has been leaked at the highest level. So that has been a concern altogether India" – IP5

"The fact remains that such a big thing [vaccine data leak] really doesn't catch the attention of people or even the-. The specialized people who are invested in that sector are the ones who really make a noise, and then after a week it just goes" - IP2

Data governance challenges involving data privacy and apparent apathy towards data leaks leads to the question of why this is the case. One of these reasons could possibly be the lack of informed consent. In the informed consent section, the rationale towards why and what informed consent means is walked through further. However, data security is something that cooperatives are able to develop and invest in as IP6 provides the example of the Uralungal Labour Contract Co-operative Society investing in cybersecurity.

However, from the secondary interviews, cooperative members seemed mindful of who they were trusting with information. This demonstrated that efforts of education regarding information sharing have been effective.

"if an unknown person approaches us for information, we do not share information with them" – Secondary Interview IP4

Value of data

As of yet, in the context of the findings of the case, there was not explicit cases where selling of data was demonstrated, however, selling of the data was not ruled out as an opportunity for cooperatives. However, the benefit provided by the cooperative is the agency of choice to do this.

"if they're organized around the business of their data, they may need to sort of sell or monetize or capitalize on the data. But here it's crucial to know is that the women themselves are making a decision on their own data" – IP1

Additionally, the aggregation of data through cooperatives provides options for cooperative members to now be connected to data, and for the data to have greater value as it is aggregated with potentially data from the whole cooperatives and other cooperatives by the federation.

"data only holds power one, when it is of a certain volume. Individualized data saving not be as useful as aggregated data so this model of women, and their cooperatives, and the federation enables this aggregation of data and enables the last person- the informal woman worker - to harness the actual power of the data." – IP1

Having access to more information, including data is enabled by cooperative networks through solidarity of members together.

"there is already a level of aggregation that happens, there is not an individual members, they are cooperative members and cooperative workers. And they have already come

together in the solidarity model and which means there's a lot of hybrid forms of information sharing across various levels in this model."– IP1

However, collecting data consistently across many cooperatives can still be a challenge as there is at times still not enough data or there is challenges in being able to derive the value from the data due to data governance issues such as data quality as discussed earlier.

"Cooperatives do not realize the gold mine of data they're already sitting on" – IP5

"some of the early responses suggest that A) still not having enough data or sometimes B) data being there but then the confidentiality of these or the neutrality of the usage does pose a challenge for now." – IP4

Time

Data collection is a time-consuming process, and there is an opportunity cost that this time spent on data collection and data a provision could be used better for something else. Learning to use digital technologies in order to be able to contribute to these services is also another thing that takes time.

Another concern with time was the "Burden of data collection" described by IP1. Data collection could be seen as another burden placed on cooperative members at the base of the pyramid with limited time that could be used instead for other activities, mainly income generating ones.

"they [women cooperative members] *need a work and income security more than anything. They need to be able to put food on the table more than anything."* – IP1

Because of the burden that data collection can place on members, there is a bigger concern that the digital divide is widening, and members are being left behind because of participation is affected. IP1 was also concerned that women are often limited with participation in discussion towards policy impacting them, such as digital inclusion. This is something that cooperatives practicing data governance aim to bridge, by providing a seat at the table in the discussion.

Equality

From the literature review, societal disparities concerning those affected by the digital divide were highlighted. However, the model of cooperatives is clear, moving towards a common objective with a common democratically chosen purpose. Thus, the code of relevancy emerged through the combined efforts of being able to use data, and who the data would be benefitting. From the interviews, it was clear that data collection was seen as having overall positive benefit for cooperatives, and their members. Nevertheless, opinions differed when it came down to the question of whether there would be someone who benefitted more than someone else.

"data is-it's really important to operate because they all [cooperative members] *are income generation mode, income generation activities. Maybe the label or the scale is different but most of the time it is income generation… Whatever data they are getting or collecting it is very- very useful for them to maintain that situation or maybe to upgrade their, their financial performance. So that kind of data might be different but the importance, I think may remain similar to most of the time."* –IP3

IP1 mentioned that in some cases when a cooperative such as a credit cooperative uses a lot of data, but the members come from a plethora of sectors then the data collected might not be directly supporting the trade they are involved in – as that is not the specific purpose of the cooperative. Thus, these members may not be directly benefitting from the data as it's not related to their trade, but the data is still crucial to collect for the operation of the cooperative.

Cooperatives are created and governed by members. Thus, cooperatives create value for their members through both economic and social incentives. Interviewees seemed to agree that within the data collection and digital sphere, cooperatives were also. With data collection schemes, it seems part of the intent is building the capacity for those with lower digital literacy levels to decrease divides rather than

"Because that's {some members benefitting more from data than others} possibly not the model of a cooperative." – IP1

"I think that's the first thing and the last thing that ultimately what do the cooperatives, what do co-ops stand for that is for the people or the for the members and for the community at large." – IP1

"I would give a direct correlation between the literacy levels, especially the digital literacy levels, to your propensity to benefit from these schemes. Higher your digital literacy levels, your ability to access or be aware of the schemes or these benefits of course- and then be able to access, does increase in a state in the country like India, wherein of course even with these benefits you are competing with several others: other members, other cooperatives to be able to access them. So, I see a direct correlation there."- IP4

"... whatever technology is introduced; it is important that is designed for the non-smartphones as well so that they can benefit from it as well – and they can receive important information from this technology." – Secondary Interview IP8

Although IP4 mentioned that having higher digital literacy may benefit access to government schemes and benefits, IP4 also mentioned that those with lower digital literacy levels are expected to benefit from programs such as the digitization of Primary Agricultural Cooperatives.

From a different level of beneficiaries, IP6 questioned whether cooperatives or the government would be the main beneficiary of the data collection projects. Of course, it is difficult to speculate, and examine the outcomes from a perspective where who benefits more would be mentioned, but this demonstrates that the question of who will benefit most from data collection and digitalization is being considered.

"Aside access, digitalization might benefit those monitoring the movement [eg. registar of cooperatives.]" – IP6

Findings Conclusion

There are many different opportunities for data usage and digitalization in cooperatives in contexts where members face the barrier of the digital divide. Cooperative members are the

ones that are expected to see a benefit, as they are making the decision and their data is the one being collected. Cooperative as organizations, are similarly seen as being some of the beneficiaries, but also those responsible for helping to organize data collection. Cooperative federations can aggregate data from multiple cooperatives to make the information garnered from data viable. Society is impacted by data governing mechanisms and the direction of where data is being collected goes. Finally, the government is responsible for administering policy and developing regulation which ultimately will shape the datafication and digitalization movement in India from a higher level.

There exists a digital divide on the access, capabilities, and outcome levels for cooperatives in India. There are differences between members digital literacy levels within cooperatives. Thus, cooperatives and cooperative federations have a few ways to bridge this divide to allow members to be able to take control of their data despite obstacles in digital literacy.

5 Discussion

From the presentation of the findings several things can be inferred:

1. The digital divides within cooperatives exist on many levels.
2. Outcomes are unclear and vary depending on the data and digital literacy of cooperatives and their members.
3. Literature does show divides within cooperatives; thus due diligence must be performed when data officers are selected
4. Informed consent is missing and requires development of digital and data literacy to foster.

In the discussion chapter of my thesis, I will add to these topics identified through the findings and examine the implications these topics may have as well as their connection to literature.

The digital divides in cooperatives

As explored in the digital divide section of the literature review, the digital divide manifests in different ways and is categorized in different ways. Wei et al. (2011) categorizes the digital divide as access, capability and outcome – from the lens of the individual, organization and global; while Van Dijk (2005) categorizes the digital divide as motivation, material, skills and usage access. Below in Figure 7, I have mapped the digital divide experiences as from findings in Wei et al. (2011) determination of digital divide categories – excluding the categorization of individual, organizational and global level that the original framework had as I felt that these were difficult to categorize with the nature of the data.

Figure 7: The Digital Divide in Cooperatives (Based on Wei et al. (2011)'s framework)

Access	Capability	Outcome
Network Coverage	Knowing how to use a smartphone	Access to government schemes
Type of device (smartphone or simple phone)	Ability to analyze data	Selling produce online
Household or personal ownership of a phone	Understanding value of data	Sharing of good governance values
	Knowing how to use phone for "productive" activities	Burden
		Potential lack of usability of data
		Accessing weather reports online to limit weather damage

From figure 7, the digital divides identified through the interviews can be examined and from this we can further break down the types of digital divides experiences. Firstly, the lack of network coverage in some of the more rural parts of India where some cooperatives are located is classified as an access level change. However, lack of network coverage is an infrastructural level division and should be acknowledged as such against other categories in the access divide, such as smartphone ownership, as the solution towards bridging this divide relies on infrastructural level investments and network coverage is something that is required to allow for a mobile phone to function to its full potential. Secondly, it was also determined that there was a difference between smart and simple phones, although it seemed that COVID-19 highlighted the need for smartphone ownership. Secondary Interview IP8 emphasized the importance of developing technologies that work on simple phones, as that is what the majority of the members have. Although a household owned a smartphone, this did not always mean that every member of the household had access to the phone. From the primary and secondary interviews, a picture of a gap between woman's access to mobile phones was identified. This corresponds to research highlighting that gender is a factor in the digital divide where more often than not women have less access to ICT (Singh 2010).

Within the literature review, Van Dijk (2005) highlighted motivation towards acquiring and using ICT devises – and mentioned that costs may be a factor inhibiting the motivation to

acquire these devices. Bijman and Wijers (2019) mentioned transaction costs as an inhibitor towards participating in democratic cooperative activity. An interviewee mentioned that data cost was "practically free" (IP6), however beyond this if mobility is required to access network coverage this could create uneven transaction costs. In the interviews a "burden" of data collection and the usability of the data to improve situations of cooperative members was discussed as a concern (IP1 & IP3). This links to the idea of participation limited by participation costs, as it highlights that choices are made based on the usability and primary financial concerns so if data and digital initiatives are not seen as important there could potentially be a hinderance to participation. However, cooperative data collection and connection can help to reduce these transaction costs by helping members both be able to access and develop ICT use capabilities.

Ownership and access to a smartphone is also interlinked to the capability divide, as through interviews and through literature it is determined that digital literacy can be enhanced through access to a phone and not just through specific education of smartphone usage. Having a phone to be able to do tasks and contact people with can allow for self-motivation to learn how to use the phone and create a situation where personal informal networks such as family, and community can be asked to assist with the usage of the phone enhancing digital literacy levels. Cooperatives help provide a network through which information is shared, both on a formal level through federations and other guiding organizations, but also on an informal level through fellow members of cooperatives.

However, even throughout the capability level of the digital divide, a plethora of different kinds of capabilities were identified. As mentioned previously, smartphone usage and access to a smartphone were identified as being connected. Moreover, being able to use a smartphone for so-called "productive work" differs from simply being able to use a smartphone. Productive work may include selling produce, accessing important information relevant to the user's occupation, or connecting with others to develop business. These types of activities can be linked to Van Dijk's (2005) skills access in the information and strategic element, while simply using the smartphone is operational skill. Even with these three skill differences existing all within the capability divide of Wei et al. (2011)'s digital divide framework, these skills are all interlinked and difficult to distinguish from each other. Being able to access information and make strategic decisions with information is reliant on the operational element of being able to use a smartphone.

With the digital divide discussed mapped out, looking at the outcomes and evaluating from the findings how cooperatives make a differential here, one of the findings that stands out is the sharing of good governance of cooperatives. Cooperatives are known to be member-based organizations working with the principles of democracy. Through the findings one implication is that the good governance cooperatives engaged in could be promoted through the digitalization of these cooperatives. This could outreach to be a greater societal benefit, demonstrating the importance of data governance structures such as data cooperatives to are based on fairness and trust (Micheli et al. 2020), potentially allowing for more people to see a way to share data but still obtain autonomy over it.

Outcomes

The outcome gap was introduced in the digital divide chapter as being the differences in outcomes from ICT use (Wei et al. 2011). In practice, it is difficult to speculate about outcomes of digitalization and data collection in cooperatives, as there are no clear metrics or clear way to measure the outcomes of digitalization and whether they would have happened without the specific initiative.

Outcomes of the digital divide are interlinked to the access and capability levels as seen in Figure 6. With better knowledge of how to use ICT "strategically" as stated by Van Dijk (2005), better outcomes could potentially be achieved. The risky outcomes investigated through the interviews included several. One of these was the concept that for cooperative members at the base of the pyramid, the data collected would not be usable and the time spent on this would be a "burden" rather than helping these cooperative members with things that are a priority for their situation such as putting food on the table and generating more income. The concept of data collection being a burden and the results not being useful, are a reminder that although bridging the digital divide is seen to enhance development (Cullen 2001), one must be critical of development policies. Who is benefiting most from the data collection may not be the cooperative, or even its members – even if there are benefits to both the cooperatives and their members. It is true that there may be more efficient ways for cooperative members to improve their immediate financial decisions. However, as mentioned by other interviewees – data schemes can greatly help connect to cooperative members to other schemes.

Cooperative's ability to provide a space for data collection and data usage with good governance principles is a very interesting concept. Thus, existing and new cooperatives forming could lead to an interesting phenomenon where the cooperative ecosystem continues to evolve, and more cooperative based data ecosystems exist. Maintaining ownership of own data, and having a trusted cooperative manage data is a benefit of cooperatives collecting and using data – as when data is collected by private entities and ownership of data is relinquished phenomena such as dispossession can occur.

Outcomes such as access to more information on for example, planting techniques or weather reports to avoid adverse weather are beneficial towards cooperative members. Outcomes of being able to obtain this kind of information reach beyond the individual themselves and project onto the rest of the cooperative and community, as there is a network of information sharing with other cooperative members and neighbours based on solidarity. Conversely, while access to new information is important, it is critical that the information used for decision making and shared to the community is accurate. Secondary interviewee IP4 explained that information was shared only after testing it themselves, to ensure accuracy or the quality of, for example, a new seed. Information accessed through Facebook or Youtube, was scrutinised in that it was not trusted as fact before trying. Moreover, information received from cooperatives, the cooperative federation, or other cooperative members was trusted thoroughly. Thus, the information shared has effects on the outcomes of ICT use.

At the current state, as stated in literature, the outcomes of ICT usage are difficult to evaluate as these outcomes are built on both access and capability levels, and as discussed previously capability levels have many nuances and variations to them. Thus, the outcomes of digitalization can be difficult to determine as many are data based and the collection and interpretation of this data is still something where capacities are being built.

Data Collection

In the inclusivity and participation in cooperatives chapter of the thesis, shifting cooperative goals and member participation were identified. Literature does show divides within cooperatives; thus, due diligence must be performed when data officers (those who are collecting data and processing it) are selected. Throughout the interviews, the findings

indicate that there is trust in the solidarity model of cooperatives and the voluntary membership of cooperatives thus benefitting members. However, the findings did also indicate that for data collection there is potential for biases to arise – meanwhile, cooperative members revealed total trust in the cooperative networks they were a part of. In literature, inclusivity in cooperatives differed (Bijman & Wijers 2019) and since data is sensitive, it is important that it is being collected with sensitivity. Since, findings did demonstrate private companies attempting to sell seeds to cooperative members and having succeeded in the past at selling seeds of inferior quality – cooperatives provide a way to protect from this through communal trust and information filtered for validity. Thus, it is crucial that those collecting and processing data are selected in a way that no third-party motives are evident.

Trade-off of between informed consent and trust-based consent

Data cooperatives are based on principles such as trust between members, the ownership of own personal data, consent towards relinquishment of said data for cooperative use with the maintenance of ownership and rights to the data at all times (Ho & Chuang 2019; Micheli et al. 2020). This means that data cooperative members have the right to withdraw their consent to their data usage at any time. Informed consent is lacking and requires development of digital and data literacy to foster.

However, interviewees expressed concerns about the levels of consent towards data related activities in India. The question was whether members could really consent to any of the data related activities if the level of digital literacy was low in the sense that there was both a lack of understanding of what information is being given away, what it will be used for, what could be the consequences, how likely are the benefits, who exactly it is shared with, how consent can be withdrawn and even what consenting to the sharing of the data fully means. This hinders principles of data governance, as consent should be informed and freely given in order to stick with principles of fairness and for members of data cooperatives to understand how they are benefitting from membership.

Consent towards the collection of data and usage of this data was established in the secondary interviews as being on a trust basis. Trust in the cooperative and other affiliated organizations was used as a basis for consent for the sharing and usage of data.

Implications

Through the findings and discussion of this case, several individual, cooperative and policy level implications can be identified. Some key implications will be highlighted in this section.

Implications for Cooperatives and Members

Firstly, from the individual level the importance of agency and self-efficacy in improving digital and data literacy is apparent. Cooperatives must be educated about the value of data, and how it can be used for their member benefits. This can enable cooperatives to provide an alternate solution towards data governance in comparison to corporations. Micheli et al. (2020) suggested that cooperatives may enable an alternative to digital capitalism, and the findings support this implication. The implications for cooperatives show that although outcomes are uncertain community, cooperative and family-based networks can help to build digital literacy and increase capabilities. However, it is important to note the information source being used and the absolute trust that members have in information provided by the cooperative and fellow members. Much work is needed before the digital divide can be fully bridged; collecting data is important as it can affect policy decisions, however governance decisions must not be taken lightly as outcomes of poor data governance as mentioned in the literature review can be drastic. Also, important to be asking the right kind of questions in order to collect the right data for the right purpose, and to be sensitive to who is collecting and processing the data.

As was suggested in literature by Weber et al. (2009), "one-size does not fit all" lin data governance. This means that despite the growing popularity of data cooperatives globally, other modes of data governance may be better for certain cooperatives. IP5 suggested the potential for data trusts as a model of structuring the data usage, demonstrating that cooperatives should think about which data structure is best for their individual situation. Thus, it is important for cooperatives to consider what kind of model best fits their individual situation.

Implications for Cooperative Federations

With all the primary interviews having been conducted with employees from cooperative federations, it is important to discuss the role of cooperative federations in relation to the research questions and findings.

Firstly, cooperative federations can promote data collection and data sharing. Although in the interviews no data sharing between cooperatives was broached, the sixth cooperative principle is cooperation between cooperatives (International Cooperative Alliance 2023). This creates an opportunity where cooperative federations could support in making the connections between cooperatives to share data when it could be beneficial, as data is only valuable when it is scalable. Proper data governance structures for this situation would have to be addressed and decided upon.

Secondly, cooperative federations can assist in providing digital and data literacy education. By promoting the benefits and educating about the risks of data collection and ICT usage, federations can promote the diversification of cooperatives. IP6 mentioned that the cooperative movement in India hadn't quite diversified as expected, thus with promotion of digital activities for cooperatives – the cooperative movement itself can progress and cooperatives can take on new activities and roles for promoting economic and social wellbeing of members. Findings found that cooperative federations are providing these services as requested by members, however, these services should be extensive to address the many barriers of the digital divide as outlined in the first section of the discussion.

Thirdly, beyond simply teaching about digital and data literacy, it is important that cooperative federations provide programs that communicate the benefits of ICT and data collection effectively. Findings indicated that cooperative members valued the benefits that could come along with ICT usage, especially monetary ones. Beyond simply communicating the benefits of collecting data and ICT, risks should be communicated as well. This would help to establish a better basis for informed consent.

Policy Level Implications

The research conducted shows that policy making decisions need to be made with care, when regarding making decisions with data for those with lower digital literacy levels. Data governance must be taken seriously to ensure that data is accurate and relevant, so better decisions can be made. From a policy level, the promotion of data cooperatives or data trusts can also create a digital landscape where citizens are more entuned with their own data and have ownership over it. However, policy should be developed in such a way that those with

the lowest access levels are not left behind (eg. simple phone vs smartphone users). This is important as gaps should not be widening.

6 Conclusion

In this final section of the thesis, I will answer my research questions as pertaining to the case – providing a summary of the findings and follow-up by examining limitations towards my work and suggesting future research topics which may be of interest.

Research Summary

RQ1: What are the trade-offs of collecting data in cooperatives?

There are many trade-offs of collecting data in cooperatives. There are many ways in which members may benefit, but there is not a clear consensus of whether this benefit will be equal for all or whether the benefit can be achieved. Firstly, collecting data in cooperatives provides a way in which cooperatives can provide an alternate data governing mechanism to the typical corporate mechanism. This mechanism seems to be based on trust, but the flipside of this is that trust between members may overwrite informed consent in sharing data at times. Secondly, members ability to make decisions is improved through data collection – however, this is only the case when members know how to analyze and use the data and when the data can be collected at a fast enough speed so that it remains current. Thirdly, the goal of cooperatives is generally that all members benefit as decisions are made democratically to ensure this, but here the risk is that if there is lack of knowledge within the cooperative on how to use the data, there may be no benefit seen, further demonstrating the affect that digital divides can have. However, if successfully implemented, data collection can create value for cooperative members, provide members ownership over their own decisions, create more entrepreneurship opportunities and even improve access to various schemes, benefit and credit.

RQ2: What is the digital divide in cooperatives?

In the case that I looked at; many different types of digital divides exist. Firstly, in some areas there is a lack of network connectivity, in other instances there is lack of access to a smartphone, while in others there is access to a smartphone but challenge in using it. From the data that I had, it seemed that being able to use ICT for "productive" or income-producing work was limited. This is to say that while there was evidence of ICT being used for "productive" things such as looking up the weather, or new farming techniques – activities

such as selling produce online seemed to be rarer. Additionally, although some members were able to use smartphones, not all were able to – and using a computer provided additional challenges. Thus, it is important to consider digital capabilities beyond access – however, it is equally important not to ignore that access to ICT can be a restricting mechanism towards digitalization of cooperatives and hinder participation of some members.

RQ3: How is data governance in cooperatives affected by the digital divide?

Data governance in cooperatives seems to be based on networks of trust – both in the form of consent being trust based, as well as information being shared through the cooperatives based on the trust members have in each other. It seemed that in general, the situation in India was described as being that people did not understand the value of their own personal data, and this extended to cooperatives. This leads to personal information being shared without a second thought, however when cooperatives are educated about this; members were careful to not share information with external parties. In cooperatives, if there is a general lack of understanding of the value of data – there can be great uncertainty regarding data governance. Additionally, data protection officers must be selected carefully to ensure trust and that they have altruistic motives due to the trust by members.

Limitations

Despite efforts to describe the phenomena seen throughout the case, there are always limitations for research. in this section I describe some of the potential limitations I identified to my case.

First of all, fewer interviews were conducted than was originally planned. Thus, the number of interviews may limit the findings and conclusions obtained throughout the thesis. However, Gaskell (2000, p.41) writes "the real purpose of qualitative research is not counting opinions or people but rather exploring the range of opinions, the different representations of the issue.". Since the purpose of qualitative research is exploring opinions instead of counting them, the number of interviews are not as crucial as the findings from those. Interviews. Additionally, it is notoriously difficult to determine at what point a point of saturation of research is reached and when no new opinions would emerge. Thus, despite limited interviews the thesis still presents different views and opinions about the topic.

Secondly, my interviews focused primarily on the cooperative federation perspective towards the topic. Thus, a different approach towards data collection could have resulted in different results and findings on the topic.

Suggestions for Future Research

Through my research, several interesting suggestions for future research emerged. I focused on early stages of data governance, such as collecting the data and analysing it for use. However, data governance consists of the whole lifecycle of the data. An avenue for future research could be studying how cooperatives with lower digital literacy levels handle redundant, old, and end of life data – and how decisions on erasing data are managed. Additionally, my research leaves over the possibility to research how data is transferred throughout the cooperative in more depth – exactly how are the informal information sharing networks cooperating.

It would also be interesting to further explore the outcome divide, this could be done by furthering my research topic sometime later when some of the projects are at a more advanced stage as at this moment responses were only speculative about outcomes.

Works Cited

Bijman, J. & Wijers, G. 2019. Exploring the inclusiveness of producer cooperatives. *Current Opinion in Environmental Sustainability,* 41**,** 74-79.

Brewer, E., Demmer, M., Du, B., Ho, M., Kam, M., Nedevschi, S., Pal, J., Patra, R., Surana, S. & Fall, K. 2005. The case for technology in developing regions. *Computer,* 38**,** 25-38.

Bühler, M. M., Calzada, I., Cane, I., Jelinek, T., Kapoor, A., Mannan, M., Mehta, S., Mookerje, V., Nübel, K., Pentland, A., Scholz, T., Siddarth, D., Tait, J., Vaitla, B. & Zhu, J. 2023. Unlocking the Power of Digital Commons: Data Cooperatives as a Pathway for Data Sovereign, Innovative and Equitable Digital Communities. *Digital,* 3**,** 146-171.

Carter, L., Liu, D. & Cantrell, C. 2020. Exploring the Intersection of the Digital Divide and Artificial Intelligence: A Hermeneutic Literature Review. *AIS Transactions on Human-Computer Interaction,* 12**,** 253-275.

Christian, B. 2020. *The Alignment Problem,* London, Atlantic Books.

Copeland, B. J. 2021. artificial intelligence. *Encyclopedia Britannica.*

Coyle, D. & Manley, A. 2022. What is the Value of Data? A review of empirical methods *The Bennett Institute for Public Policy* [Online].

Crawford, K. 2021. *Atlas of AI,* New Haven and London, Yale University Press.

Creswell, J. W. (2009). Research design: Qualitative, quantitative, and mixed methods approaches (3rd ed.). Sage Publications, Inc.

Cullen, R. 2001. Addressing the digital divide. *Online Information Review,* 25**,** 311-320.

Decuir-Gunby, J. T., Marshall, P. L. & Mcculloch, A. W. 2011. Developing and Using a Codebook for the Analysis of Interview Data: An Example from a Professional Development Research Project. *Field Methods,* 23**,** 136-155.

Dohmwirth, C. & Hanisch, M. 2019. Women's active participation and gender homogeneity: Evidence from the South Indian dairy cooperative sector. *Journal of Rural Studies,* 72**,** 125-135.

Gray, C. 2021. Data Dispossession. *The Sociological Review Magazine.*

Hanbal, R. D., Prakash, A. & Srinivasan, J. 2020. Who drives data in data-driven governance? The politics of data production in India's livelihood program. *Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance.* Athens, Greece: Association for Computing Machinery.

Hardjono, T. & Pentland, A. 2019. *Data Cooperatives: Towards a Foundation for Decentralized Personal Data Management.*

Harvey, D. 2004. The "New Imperialism": Accumulation by Dispossession. *Actuel Marx,* 35**,** 71-90.

Hassan, R. 2020. The Economy of Digitality Limitless Virtual Space and Network Time. *The Condition of Digitality.* University of Westminster Press.

Ho, C.-H. & Chuang, T.-R. 2019. GOVERNANCE OF COMMUNAL DATA SHARING. *Good Data.* Institute of Network Cultures.

International Cooperative Alliance. 2018. *Cooperative identity, values & principles* [Online]. International Cooperative Alliance. Available: https://www.ica.coop/en/whats-co-op/co-operative-identity-values-principles [Accessed 2023].

Ito, J., Bao, Z. & Su, Q. 2012. Distributional effects of agricultural cooperatives in China: Exclusion of smallholders and potential gains on participation. *Food Policy,* 37**,** 700-709.

Jakku, E., Taylor, B., Fleming, A., Mason, C., Fielke, S., Sounness, C. & Thorburn, P. 2019. "If they don't tell us what they do with it, why would we trust them?" Trust, transparency and benefit-sharing in Smart Farming. *NJAS - Wageningen Journal of Life Sciences,* 90-91**,** 100285.

Jamil, S. 2021. From digital divide to digital inclusion: Challenges for wide-ranging digitalization in Pakistan. *Telecommunications Policy,* 45**,** 102206.

Janssen, M., Brous, P., Estevez, E., Barbosa, L. S. & Janowski, T. 2020. Data governance: Organizing data for trustworthy Artificial Intelligence. *Government Information Quarterly,* 37**,** 101493.

KPMG. 2022. *Digital Personal Data Protection Bill* [Online]. KPMG. Available: https://kpmg.com/in/en/home/insights/2022/12/privacy-digital-personal-data-protection-bill-2022.html [Accessed].

Kumar, A., Saroj, S., Joshi, P. K. & Takeshima, H. 2018. Does cooperative membership improve household welfare? Evidence from a panel data analysis of smallholder dairy farmers in Bihar, India. *Food Policy,* 75**,** 24-36.

Lythreatis, S., Singh, S. K. & El-Kassar, A.-N. 2022. The digital divide: A review and future research agenda. *Technological Forecasting and Social Change,* 175**,** 121359.

Mehrabi, Z., Mcdowell, M. J., Ricciardi, V., Levers, C., Martinez, J. D., Mehrabi, N., Wittman, H., Ramankutty, N. & Jarvis, A. 2021. The global divide in data-driven farming. *Nature Sustainability,* 4**,** 154-160.

MEITY 2022. THE DIGITAL PERSONAL DATA PROTECTION BILL, 2022. *In:* MEITY (ed.).

Micheli, M., Ponti, M., Craglia, M. & Berti Suman, A. 2020. Emerging models of data governance in the age of datafication. *Big Data & Society,* 7**,** 2053951720948087.

Miller, K. 2021. *Radical Proposal: Data Cooperatives Could Give Us More Power Over Our Data* [Online]. Stanford University Human-Centered Artificial Intelligence. Available: https://hai.stanford.edu/news/radical-proposal-data-cooperatives-could-give-us-more-power-over-our-data [Accessed March 7th, 2022].

Mulgan, G. & Straub, V. 2019. *The new ecosystem of trust* [Online]. Nesta. Available: https://www.nesta.org.uk/blog/new-ecosystem-trust/ [Accessed 2023].

Österberg, P. & Nilsson, J. 2009. Members' Perception of Their Participation in the Governance of Cooperatives: The Key to Trust and Commitment in Agricultural Cooperatives. *Agribusiness,* 25**,** 181-197.

Patton, M. Q. 2015. *Qualitative Research & Evaluation Methods,* Thousand Oaks, California, SAGE Publications.

Perez, C. C. 2019. *Invisible Women,* London, Vintage.

Puusa, A., Mönkkönen, K. & Varis, A. 2013. Mission lost? Dilemmatic dual nature of co-operatives. *Journal of Co-operative Organization and Management,* 1**,** 6-14.

Rajam, V., Reddy, A. B. & Banerjee, S. 2021. Explaining caste-based digital divide in India. *Telematics and Informatics,* 65**,** 101719.

Ranis, P. 2016. *Cooperatives confront capitalism : challenging the neoliberal economy,* London, England, Zed Books.

Rhodes, R. a. W. 1996. The New Governance: Governing without Government1. *Political Studies,* 44**,** 652-667.

Riggins, F. & Dewan, S. 2005. The Digital Divide: Current and Future Research Directions. *J. AIS,* 6.

Ruhanen, L., Scott, N., Ritchie, B. & Tkaczynski, A. 2010. Governance: a review and synthesis of the literature. *Tourism Review,* 65**,** 4-16.

Scheerder, A., Van Deursen, A. & Van Dijk, J. 2017. Determinants of Internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide. *Telematics and Informatics,* 34**,** 1607-1624.

Scholz, T. & Calzada, I. 2021. *Data Cooperatives for Pandemic Times*.

Shah, S. I. H., Abdulaal, A. & Peristeras, V. 2022. Data divide in digital trade, and its impacts on the digital economy: A literature review. *Proceedings of the 15th International Conference on Theory and Practice of Electronic Governance.* Guimarães, Portugal: Association for Computing Machinery.

Singh, S. 2010. Digital Divide in India: Measurement, Determinants and Policy for Addressing the Challenges in Bridging the Digital Divide. *IJIDE,* 1**,** 1-24.

Singh, S. 2012. New markets for smallholders in India: Exclusion, policy and mechanisms. 47, 95-105.

Stoker, G. 1998. Governance as theory: five propositions. *International Social Science Journal,* 50, 17-28.

Tewathia, N., Kamath, A. & Ilavarasan, P. V. 2020. Social inequalities, fundamental inequities, and recurring of the digital divide: Insights from India. *Technology in Society,* 61, 101251.

Thatcher, J., O'Sullivan, D. & Mahmoudi, D. 2016. Data colonialism through accumulation by dispossession: New metaphors for daily data. *Environment and Planning D: Society and Space,* 34, 990-1006.

Van Deursen, A. J. & Van Dijk, J. A. 2019. The first-level digital divide shifts from inequalities in physical access to inequalities in material access. *New Media & Society,* 21, 354-375.

Van Dijk, J. 2005. The Deepening Divide: Inequality in the Information Society. Thousand Oaks, California.

Van Dijk, J. a. G. M. 2019. *The Digital Divide*. Polity Press.

Weber, K., Otto, B. & Österle, H. 2009. One Size Does Not Fit All---A Contingency Approach to Data Governance. *ACM Journal of Data and Information Quality,* 1, Article 4.

Wei, K.-K., Teo, H.-H., Chan, H. C. & Tan, B. C. Y. 2011. Conceptualizing and Testing a Social Cognitive Model of the Digital Divide. *Information Systems Research,* 22, 170-187.

Wu, D., Verhulst, S. G., Pentland, A., Avila, T., Finch, K. & Gupta, A. 2021. How data governance technologies can democratize data sharing for community well-being. *Data & Policy,* 3, e14.

Zins, C. 2007. Conceptual approaches for defining data, information, and knowledge. Journal of the American Society for Information Science and Technology, 58, 479-493.

Zuboff, S. 2019. *The Age of Surveillance Capitalism,* London, Profile Books.

Appendix 1

1. How do differences in digital literacy in the cooperative setting impact participation in data collection for AI? Can you give an example?

    a. ALT: Can you think of a situation where differences in digital literacy have directly impacted participation in data collection for AI?

2. How can members with low digital literacy benefit from the data they are providing?

    a. Can you give an example where they have?

    b. How do you see this changing in the future?

3. Can you specifically ensure that data used is specifically benefitting those who have lower digital literacy or less access to technology?

    a. How do you ensure it?

    a. If no, what are some examples of times data has not benefitted these groups? How can we learn from these situations to ensure that data does benefit those with less access to technology?

4. How do you ensure the data you collect is relevant?

    a. Can you give a specific example?

5. Might there be situations when the use of data benefits some cooperative members more than others?

    a. How do you think these trade-offs might be managed?

Potential follow ups:

    i. Is it OK as long as everyone is at least benefitting a little?

    ii. Is it OK as long as people who are the 'worst off' are benefitting a little more?

6. Are there risks of commodification of data?

    i. Are there risks of externals from cooperatives profiting from the data collection?

    ii. Do you consider this a risk?

    iii. If there are risks of data commodification, could commodification benefit members?

Wrap up questions: When it comes to the use of data in cooperatives, what do worry about? /What problems keep you up at night?

Anything else that you want to bring up, that I didn't ask in the interview?

Appendix 2

Codebook

Group	Code	Code Description	Example
Data Governance	Data quality	This code is used when data quality is mentioned	"biases do pour it in and there's sometimes the participation is very much mobilized by such biases" – IP4
	Access	This code is used to categorize when access for cooperatives or members is mentioned. This code excludes access in terms of the first level digital divide.	"actually monitoring all kinds of welfare schemes or mechanisms by which the government can reach these last-mile entities last-mile villages last-mile pockets in the country so uh data on the gender ratio or the literacy rates in the in a certain village or the presence of youth leaders or you know policymakers in that part of the country." – IP4
	Relevancy	This code is mentioned when the relevance or applicability of data is mentioned	"because we already have some difficulties in collecting data, we only collects stuff that we know is going to be useful, not just you know collect it and then try to find a use for it"- IP2
Challenges	Time	This code is used when challenges regarding time are mentioned.	"once we do collect it maybe it's become redundant right because for businesses you need accurate data really fast" – IP1
	Interpretation	This code is used when there is struggles to analyze and create interpretations of the data.	Use this code when there is exchange between cooperatives (or lack of)
	Resources	This code is used when the resource intensity of data collection is brought up.	Use when individuals make decisions for themselves.
	Consent	Use code when challenges with consent to data collection/storage/usage are identified.	"Then thirdly you know inform the consent so all these I would say would be the problems of any cooperative from using data or collecting data in the first place and of course the cooperatives themselves not entirely being aware of the potential of such data."- IP5

Opportunities	Solidarity	When there is a sense of solidarity/unity between members.	"the federation enables this aggregation of data and enables the last person- the informal woman worker to harness the actual power of the data" – IP1
	Connection	Use code when there is an increase in connectivity.	"I think starting from online meetings, filing of-of you know notes, annual general body meetings, and you know all statutory requirements, has a basic requirement. I think a lot of that has been happening online." – IP5
	Entrepreneurship	When entrepreneurial purposes are identified.	"So whenever we want to learn a new application, we ask others to teach us and then we teach other people too." – Secondary Interview IP8
	Learning	When there is some learning or training.	"specialized services and the federation provides like capacity building" – IP1
	Outreach	When more are reached through cooperative based data collection projects.	"reach out to you know the last persons in the in the village or in the countryside who are usually not having good access to digital, you know, tools." – IP4
	Decision-making	When data is used to make decisions.	"I think they [cooperatives] will benefit a lot from having more accurate effective data to make these decisions, which of course in the long run can help in providing the right kind of environment to the members of these cooperatives: socially, environmentally, economically." - IP4
Network	Cooperative-based	This code signifies the strong personal ties cooperative members have with each other.	"The Mandli whatsapp group also helps us a lot, by giving us advice on what kind of seeds and processes to use depending on what kind of land and soil we have, what kind of harvest we might get" – Secondary Interview IP8
	Family	This code signifies the strong personal ties with family	"Our children are well versed with technology now, so they help us out with these apps if

			we get stuck anywhere." – Secondary Interview IP1
	Government	This code is when the government or a regulatory framework is directly mentioned.	"Firstly, you have the Indian government which is you know in mission mode, which is trying to digitalize rural cooperative banks at the at the very primary level" – IP6
	Trust	This code is mentioned when trust is either explicitly mentioned or it is implicit.	"All the data used for the benefits of the members only and they have a high trust in sharing with the cooperatives so they do share." – IP2
	Agency	When cooperative members have a freedom to change their ways of doing things from self-motivation.	"I mean finally ownership because like I said owning the data is and being able to make decisions about and using your data is crucial" – IP1
Data commodification	Marketing	This code signifies when a product is being sold/marketed to cooperatives/their members.	"First private companies used to come and sell us seeds for an expensive price, and we were not informed how to use the seed as such, and so we ended up spending a lot without any great results." – Secondary Interview IP7
	Data-selling	This code is used when selling data is mentioned.	"using this data for informed decision making even for monetizing that would be something that existing cooperatives can do" – IP5
Digital Divide	First-Level Digital Divide	The First-Level Digital divide or "access" divide as defined by Wei et al. (2011)	"low internet connectivity issues often become a barrier" – IP2
	Second-Level Digital Divide	The Second-Level Digital divide or "capability" divide as defined by Wei et al. (2011)	"they don't know how to analysis the data and what the data says all about" – IP2
	Third-Level Digital Divide	The Third-Level Digital divide or "outcome" divide as defined by Wei et al. (2011)	"Higher your digital literacy levels your ability to access or be aware of the schemes or these benefits" – IP4