

Re-imagining the Platform Firm: Lessons and Design Blueprints from SEWA's Data Cooperative Experiment

Ranjitha Kumar¹, Salonie Muralidhara Hiriyur²

¹IT for Change, India; ²SEWA Cooperative Federation and Institute of Development Studies, IDS
ranjitha@itforchange.net

**Paper prepared for presentation at the
"8th Conference of the Regulating for Decent Work Network"
on Ensuring decent work in times of uncertainty
at the International Labour Office Geneva, Switzerland
10-12 July 2023**

Abstract

The dynamics of capital and labour undergo significant changes as various production processes across sectors are being brought into the fold of digitalisation, particularly at the cost of informal women workers in agriculture. The growing trend of agricultural digitalisation and datafication, particularly in the Global South, has solidified existing class precarities, and has modified land relations among several marginalised wage workers and farmers. With land sovereignty being increasingly pegged to data sovereignty, the expropriation of data by a handful of Big Tech corporations has immense consequences on existing tenure rights and relationships. This research study, conducted in collaboration with SEWA Cooperative Federation, investigates the following questions in response to the threat of datafication exacerbating existing precarities of women informal agricultural workers in the Global South: (1) What kind of alternative data systems can be created by adopting a base-up mapping of women's data needs into SEWA's process of building a worker-designed and led data cooperative and (2) how can this process of trust-based co-design enable the creation of a data cooperative that circumvents the extractive paradigms of data capitalism, with the infrastructure and governance principles to resist the extension of Big Tech into Global South food systems. In this paper, we demonstrate the need for developing base-up models of data cooperatives as an alternative to top-down data cooperatives, the latter often centralised and non-federated, built to enhance data sharing and monetisation and at the mercy of digital

monopolies. Through a Rapid Ethnographic Approach, our Participatory Action Research adopts a unique consultative co-design process ((N = 60 women farmers), where we go on to propose a federated techno-institutional model that is pegged on SEWA's existing, unique hyperlocal Farmer Facilitation Centres (FFC) and Mandli systems. The study highlights the need for techno-design to be informed by data rights principles, foregrounded by political economy narratives that critically appraise digital systems within data capitalism. It also highlights the importance of developing grassroots research methodologies to implement the concept of digital commons as a viable means of resistance against the proliferation of datafication in the Global South. The techno-design blueprints emerging from this collaboration are significant additions to the body of work on executing platform cooperative models.

Keywords: Data Cooperative, Rapid Ethnography, Techno-Institutional Model, Data Rights, Women Agricultural Workers

Copyright 2023 by author(s). All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Introduction

Data Cooperatives and The Economic Value of Data

Against the phenomenon of digitalisation - the integration of digital technologies within the value chains of various economic sectors, often producing new forms of revenue - and the subsequent entry of all information as data points into the ontological paradigm of datafication (Mejias & Couldry 2019; Sadowski 2019), there is a growing need for the recognition of data as an economic resource, and its governance through a community rights framework (Singh & Gurusurthy, 2021). Additionally, the value data possesses emerges from its pooled and processed form, and hence, demands both a regulatory and epistemological framework that views data within digital capitalism¹ as a community resource.

The imperative of ensuring economic gains made from pooling of data are enjoyed both by the individual and the community has been recognised by several scholars who critically examine the growth of platform capitalism (Papadimitropoulos, 2021). One such narrative of challenging the sharing economy comes from the school of platform cooperativism (Scholz, 2016); emerging as a worker-led and owned alternative to platform capitalism, platform cooperativism aims to develop an iterative version of the cooperative movement - historically embedded in the Global South for several decades - as a means to transform the exploitative economic dialectics of today's platform economy.

The rapid expansion of digitalisation has also resulted in the emergence of data infrastructures in various other economic sectors, which are now transforming these value chains inside out (Fraser, 2019). The digitalisation of food systems through the confluence of the top players in agriculture, technology, and finance has resulted in a tight control of the agricultural chain by big data and transnational corporations, often at the detriment of the environment and agriculturalists (ETC Group, 2022). As the logics of the data capitalism rearrange the economic and social relations of sectors and transform informational databases into enclosures for profits (Gurusurthy et al., 2022b), there is concern over the data rights of workers from these sectors, as well as the expropriation of vital, community-owned data on agricultural practices by large agriculture-technology (ag-tech) and multinational corporations.

In parallel to platform cooperativism, interest groups and organisations are now exploring bottom-up data infrastructures (Keller, 2021) as a means to redistribute economic gains and protect data from

¹ In this paper, the term data and digital capitalism have been used interchangeably; it is however important to note that these terms have emerged from different socio-political debates at parallel points in history.

corporate misappropriation. A data cooperative is formed by individuals or economic actors pooling, organising, and analysing data to generate high-quality aggregated data for revenue redistribution (Mehta et al. 2022; Gurumurthy & Chami 2022a). Additionally, data cooperatives also facilitate the addition of digital infrastructure; in the case of an agricultural data cooperative, this service can include real-time data on weather forecasting, market rates and instant message services to enable clearing-sharing of agricultural techniques. Data cooperatives come under the larger paradigm of data stewards - data intermediaries who facilitate or hold consent and decision-making on behalf of a group of data users (Manohar et al., 2020) - with the aim to actualise the collective rights of data (Singh & Gurumurthy, 2021).

However, there is a lacunae of ground-up approaches to building data cooperatives. Data cooperatives under the open data paradigm - building data cooperatives that maximise centralisation and monetization of data - often replicate existing data systems (Massey et al., 2023) within the regime of datafication. These systems are also rarely accompanied by a rights-based approach to encoding the rights of data generators and workers; often making them susceptible to incursions by Big Tech and Big Ag companies, either through inevitable purchasing of collective data, or by becoming key infrastructural players within the data architecture of such cooperatives, gaining an amount tremendous amount of pooled information from the process (Amazon Web Services, for example, are one of the few ways in which communities can access server space in today's digital economy) (Pentland & Hardjono, 2020). Additionally, most data cooperatives face the daunting task of seamlessly incorporating the "data layer" into the economic and social relations of the cooperative.

As the juggernaut of data capitalism and financialisation only saw newer heights through the COVID-19 pandemic (Fernandez et al., 2020), there is an urgent need to revisit the idea of participatory approaches to designing digital interventions (Latonero & Aneja, 2021), in addition to finding avenues to regain and redistribute the value of pooled data. Top-down approaches often result in systems that do not serve the purpose of interest groups, continuing to aid the expropriation of data's economic value (Sandoval, 2020). This study aims to co-create the building blocks of a federated data cooperative by expanding on two fronts: (1) a data rights-approach to enshrining governing principles of a data cooperative and (2) creating a confluence of theory-driven methodologies of ethnographies with participatory co-design which can ultimately enable agentic co-creation of technologies, serving the needs of data communities.

Embedding Trust in Data Stewardships: Co-Creation of a Techno-Institutional Framework by SEWA Cooperative Federation and IT for Change

SEWA Cooperative Federation is one of the largest labour-managed cooperatives in India with over 300,000 active women members. SEWA Cooperative Federation's² focus areas include meaningful digital inclusion of informal women workers and their cooperatives/social enterprises. In the agriculture sector, the Federation has been working with cooperatives to enable the use of digital technologies for communication, monitoring and learning, and business development. In the current project, multiple research partners are currently engaged in proposing various moving parts of the data cooperative, the longer-term goal being to co-creating a blueprint of a sustainable, scalable women-owned cooperatives, which includes the goal of building the capacities of women workers and leaders to own, manage and use data and data-based systems to improve their own work, income security and social security.

As a part of IT for Change's 'Re-wiring India's Digitalising Economy for Women's Rights and Well-being: an Action-oriented Knowledge Intervention'³ SEWA and IT for Change (ITfC) have embarked on a knowledge partnership that explores the construction of inclusive data models, driven by the need for a base-up approach to co-design, placed well within the critical explorations of data within digital capitalism.

Through this research, ITfC and SEWA aim to propose a techno-institutional framework built through Participatory Action Research. In their initial quantitative assessment of the field (Kumar et al., 2022), trust in data and data systems emerged as a significant barrier of entry for women farmers. Gender-mediated ownership and access to technology, in addition to the "double burden" of (informal) work and domestic work (Bhattacharya 2017; Chakraborty 2021) has significantly impacted women agricultural workers, who often contribute to the most labour-intensive processes within agriculture, but are often not classified as farmers (Pattnaik & Lahiri-Dutt, 2022).

Moving beyond the economic theorisation of trust and its historical manifestation (Borgen, 2002) within cooperatives, the research projects aims to embed trust in the "data layer" through the processes of a ground-up consultation with farmers of the cooperative, with the purpose to reverse

² We use the term SEWA and SEWA Cooperative Federation interchangeably to specifically refer to the Cooperative Federation, which is part of the larger SEWA movement

³ Re-wiring India's Digitalising Economy for Women's Rights and Well-being: an Action-oriented Knowledge Intervention is an IT for Change project, funded by Friedrich-Ebert-Stiftung (FES), with support from the European Commission (EU). This is a five-year (2020-2024) project on Gender and the Digital Economy, with the aim to ensure that the benefits of a data-driven and digitised economy accrue to women. More details about IT for Change's project can be found at

<https://itforchange.net/digital-economy-womens-rights-wellbeing-india>.

the top-down approach to creating data communities. This alternative, agentic approach can assist in the formation of pluralistic, federated data cooperatives that function interdependently, and are equipped with robust data governance frameworks, protecting the interest of women agricultural workers as well as offer actionable pathways for achieving redistribution of economic gains from aggregated data.

Research Methodology

The goal of this research project is to synthesise a base-up process of co-creating a data cooperative, that enables its various actors - data generators, aggregators and processors - to democratically contribute to its formation. The historicity of SEWA emerging out a union movement and its success as a cooperative provides the project with ample means of democratising collective economic value; through this study, we aim to embed those values, avenues and lessons into a data cooperative that serves the economic and social needs of women agricultural workers, effectively addressing the gender-rural digital divide.

The following are the primary research questions of the study:

- 1) What kind of alternative data systems can be created by adopting a base-up mapping of women's data needs into SEWA's process of building a worker-designed and led data cooperative?
- 2) How can a process of trust-based co-design enable the creation of a data cooperative that circumvents the extractive paradigms of data capitalism, with the infrastructure and governance principles equipping the system to resist the extension of Big Tech into Global South food systems?

Methodology

The study adopts two key research methods: Rapid Ethnography (RE) that imbibes a Grounded-Theory Approach, contextualised by the principles of a Participatory Action Research (PAR). Rapid Ethnographies are suitable for short-term, intensive engagement within the field. The collaborative research team consisting of two members from IT for Change accompanied by a SEWA member spent one week across two Mandlis in Gujarat, India: Megha Mandli in Tapi and Kheda Mandli in Kheda. The goal of this methodology is to implement ground-up techniques that are core to ethnographic enquiries in harmony with the analytical frameworks of PARs - a qualitative research methodology that demands the collaboration between participants and researchers in order to exact actionable change. The PAR process, which is alternatively termed as co-design/consultative process involves comprehensively incorporating all inputs from the participants of the study into the final

output, where any proposed results are built in collaboration on the field and not independently post-hoc. Hence, a grounded theory lens was crucial to this process; any and all research instruments underwent significant modifications after each research day, informed by participants' inputs, conversations with SEWA leaders and through a thorough reflexivity process of co-creating field notes and journaling.

Methods

The PAR co-design process took place through a series of Focus Group Discussions (FGDs), each group consisting of 8-10 members belonging to the same interest group. Most farmers interviewed were marginal land owners, and often also work as agricultural wage workers in other farms. Across Megha and Kheda, the research team visited 2 Adivasi villages which consisted of workers who were part of the Sardar Sarovar dam-resettlement and rehabilitation programme. The other villages were caste-diverse but almost always consisted of farmers who were Dalit, Bahujan or Adivasi.

Table 1: Research Methods

Mandli	Total FGDs	Interest Groups and FGD Location	Questions and Themes
Megha Mandli	4	Farmers - Village 1	Part 1: Dimension mapping the political economy of trust in SEWA - capturing existing Informational flows and centralisations within SEWA. Part 2: What are the norms, attitudes, perceptions and usage data across various member-types of SEWA? Part 3: Within the frameworks of productivity enhancement, how can we harness the advantages of a data cooperative to aid caste and class marginalised women farmers, often practising subsistence agriculture?
		Farmers - Village 2	
		Aagewans - Mandli	
		Board Members - Mendli	
Kheda Mandli	4	Farmers - Village 1	Part 3: Within the frameworks of productivity enhancement, how can we harness the advantages of a data cooperative to aid caste and class marginalised women farmers, often practising subsistence agriculture?
		Farmers - Village 2	
		Male Farmers - Mandli	
		Aagewans and Board Members - Mandli	

An average total of sixty women across the two districts participated in the FGDs. Each FGD was two hours long - oftentimes, questions asked had to go through a translational chain from English to Hindi/Gujarati, and eventually to local dialects and vice-versa, facilitated by the SEWA member present during the process.

Results

The Structure of SEWA Cooperative Federation

As a worker-led movement, SEWA has a decentralised internal structure with varying degrees of management and oversight. It is important to know due to its highly federated model, responsibilities that are assigned to leadership positions differ at various levels, depending on the needs of each community and cooperative, as well as decisions made during committee meetings.

The interviews with women across Mandlis revealed that there's an inherent understanding of data, its ability to be pooled and the advantages one can reap from data that has been collected, processed and brought into the helm of becoming inferential. The appreciation for data is seen across various types of digital users, as well as those who have varying accessibility to ICTs and smartphones. Additionally, this understanding permeates several class, caste, location and positional boundaries within SEWA. Existing data - collected through manual techniques - have already shaped the functioning of the cooperative, indicating that there exists specific data cycles within SEWA that must be effectively incorporated into an emerging digital paradigm.

Leadership Level: Women Members of SEWA-supported cooperatives

Women farmers access membership to SEWA Union through an annual fee of ten rupees, and can become shareholders of a trade-based cooperative (here, agriculture) through purchase of shares, as defined in the cooperative's bye-laws.

Within the cooperative model, there are layers of decision-making, culminating at the cooperative Board. A sub-committee meets at the village-level, and may include several villages, based on the density of members in the specific area. The meetings may also include farmers who are not cooperative members, and also often includes men. Social relations between the members frequently overlap particularly during panchayat meetings.

Leadership Level: Aagewan

Members can graduate into 'Aagewans' (translates to: woman who comes forward), and these are women farmers in the village nominated as community leaders, and represent the farmers from the village at various other touch points such as board meetings, panchayat meetings, meetings with other external stakeholders and researchers who work with the cooperative and/or the Federation. It is important to note that Aagewans are nominated through a deliberation process held at the community level; there isn't a traditionally understood election process where leaders are chosen anonymously through the community.

During the consultation process with SEWA farmers, some women farmers mentioned that they would prefer that Aagewans are chosen through a formal election process and on a rotational basis, in addition to the existing deliberation process - it was revealed that caste and class factors play a role in leadership determination. Approaching this through an ontological lens that presupposes ballot elections as fairer would be a faulty assessment of the situation; nonetheless, elucidating on SEWA's leadership practices are key to understanding how data collectivism through the data cooperative can not only enable the redistribution of economic gains of aggregation, but can also effectively enhance existing cooperative practices. What are the ways in which the data cooperative can alleviate concerns of collectivism and unionisation as it exists within a rural Indian political economy are some of the ancillary techno-design questions to expand.

Leadership Level: Sankalit Sathi

Sankalit Sathis (translation: those who bring together) lead and coordinate with Aagewans within a group of nearby villages. They are paid by their cooperative. The primary goal of Sankalit Sathis is to understand the goals and demands of the members of SEWA, and act as conduits to translating this information to board members. Oftentimes, they represent the concerns that have emerged during meetings with Aagewans and women members of the cooperative at board meetings and collaborate across parts of the Federation to make decisions.

Leadership Level: Kaarobari

Kaarobaris (Translation: board members) are women who represent their village(s) at the cooperative Board are nominated and elected through a democratic process by the cooperative. These leaders work to fulfil the requirements of the cooperative. They take strategic decisions, represent their cooperatives at various fora and to relevant local, national, international stakeholders, develop and monitor business plans. They also work directly in business activities. This could include the purchasing of seeds, taking stock of what is needed at a cooperative level, assessing needs of hyperlocal centres and subsequent villages. It is crucial to note that some Aagewans are also Kaarobaris; these overlap in leadership structures and responsibilities are response to the needs of individual villages.

Structural Level 1: Villages

At the village level, women enrol into the cooperative and attend regular committee meetings. Aagewans are often nominated through these committees and are also women farmers from the same village or neighbouring villages.

Structural Level 2: Farmer Facilitation Centres

A Farmer Facilitation Centre is a hyperlocal centre that offers agricultural services and products to a handful of nearby villages. Often headed by an Aagewan or a board member, FFCs are touch points for women to approach for specific input-related services (such as access to seeds, fertilisers, etc) as well as farming advice. The cooperative may also employ a staff person to manage the FFCs, and coordinate with Mandlis to ensure FFCs are appropriately stocked and have up to date information on market price/APMC prices, availability of local merchants, etc.

Structural Level 3: Mandli/Cooperatives and the Cooperative Federation

Mandlis (or cooperatives) are members of the SEWA Cooperative Federation, which itself is a secondary-level cooperative⁴. Through their Board, Mandlis make several independent decisions for its constituents, and are structured in a manner that allows them to function independently, and make agile decisions based on the context of individual villages. These are membership-based entities, and in the case of SEWA, are owned and managed completely by women farmers.

The Federation too has a Board of Directors, who are nominated and elected democratically for a definite term, from the pool of all Board members or all member cooperatives. Their election happens at the Annual General Meeting of the Federation, and is based on a sectoral representation decided based on total membership of the sector in the Federation. For example, the agriculture sector is the largest constituent of the Federation and holds a proportional number of seats at the Federation Board.

⁴ A secondary-level cooperative here would mean that SEWA Cooperative Federation's members are primary-cooperatives, and not individual women workers.

SEWA's Value Addition to Farming, Farming Lives and Livelihoods

Women farmers interviewed understand SEWA as not just a cooperative of women, but of women farmers/workers who have come together for an economic purpose; i.e. productivity enhancement from their own farmers, and the maximise the benefits of their own labour and agricultural practices. On that end, the following is a mapping of SEWA's value-add to the agricultural value chain, particularly those that benefit women farmers and producers at the lowest rung of the agricultural production process.

Table 2

Agriculture			Agricultural-allied Services		Non-agricultural Services	
Input-related materials	Input-related services	Market information	SEWA as a market	Livelihood schemes	Healthcare	Grievance redressal
<ul style="list-style-type: none"> - Seed procurement for farmers, availability of seeds at a discounted rate depending on the need of the village - Fertilisers, manure, and farming-related materials - Some FFCs have been given machinery by the cooperative for processing goods, such as small mills that SEWA farmers 	<ul style="list-style-type: none"> - Information on seeds, variety of seeds, cost of seed procurement as well as seasonal-related information on what seeds can be grown. - Information on cropping, harvesting and planting techniques, advice on trouble-shooting through FFCs. - FFC at each Mandli also advise on new and improved tools to 	<ul style="list-style-type: none"> - FFCs provide daily market rates to women in person. However, the location of villages greatly determines accessibility to this information. Since Naranpura - the first village where co-design took place - is closer to the main FFC, most women reported visiting the FFC for information, as well as to gain knowledge on input services and other agricultural practices. However, the consultation process at Dhupi revealed 	<ul style="list-style-type: none"> - Though minimal in capacity, SEWA also purchases select crops from farmers through the Megha Mandli. - Certain crops, such as specific millets, red and black rice are occasionally purchased by the Mandli at a marginally better rate than those offered by both merchants and APMCs. SEWA and Megha Mandli make the decision to buy certain crops based on 	<ul style="list-style-type: none"> Beyond agricultural input and output support services, SEWA also facilitates agricultural-related scheme enrollment services. FFCs/Aagewans are often information points who help women determine schemes to choose, and the process for enrollment. Aagewans will also take forth their complaints and feedback to panchayat meetings, and will 	<ul style="list-style-type: none"> Particularly in the context of the pandemic, women reported that SEWA facilitated access to several healthcare services. These include provision of medicines to remote villages (through the FFCs and Aagewans), aiding access to hospitals and healthcare centres, and providing timely information on the pandemic, along with healthcare advice. 	<ul style="list-style-type: none"> A primary function of SEWA's structure is to facilitate environments of participatory governance and grievance redressal. Regular meetings with Aagewans are held at each village level, when women come together to address both agricultural and non-agricultural issues. Women also expressed how there's limited scope for grievance redressal

<p>can use if needed, instead of reaching out to merchants</p>	<p>purchase, as well as facilitated purchases of the above.</p>	<p>challenges that were not faced by Naranpura; since Dhupi was further away from the FFC, women rarely have the time to visit the FFC, and do not gain real time information.</p>	<p>demand constraints in the market, as well as how many farmers grow this crop; if specific millets do not have much demand in the market, SEWA offers to buy these millets from small and marginal farmers at a favourable rate.</p>	<p>facilitate information sharing between women farmers across multiple villages and the panchayats on the matter of challenges faced by women during enrolling into schemes.</p>	<p>Through the federation ecosystem, the healthcare cooperative - Lok Swasthya Mandali - frequently engages with the agriculture cooperatives in this regard.</p>	<p>within SEWA as the challenges they face are more systematic and infrastructural; panchayats, rather than SEWA, tend to make decisions on the above (such as daily wage rate).</p>
--	---	--	--	---	---	--

Data Needs of Women Farmers in SEWA

The primary research question of this project was to ascertain the kind of data women would like to receive from the cooperative. Though initial consultations by the research team focussed on agricultural data, it soon emerged that women farmers are looking for more comprehensive data analytics from SEWA, along with aggregated and non-aggregated data on all aspects of livelihood, and not just farm-related information.

Data Services

Women's imagination of a data cooperative was often tied down to the idea of a phone application (app) that one can access through their phones. One theorisation of this popular understanding - seen across the four villages - was due to the limits and liberties of translation, as data systems often do not have a regional language counterpart, and are often translated to as "apps". Previous co-design and research processes conducted by other partners have also instilled a specific imagination of how the data cooperative will function and federate itself, which influenced this specific consultative process.

Though access to aggregated/pooled data materialised as a prime concern, the imagination of an application also resulted in women expecting certain data services independent of aggregation from the cooperative. Though SEWA is in the process of determining the feasibility of building such an application, it is pertinent to capture the techno-design requisites of women farmers. The following are some common non-aggregated data services that women farmers would like to receive through a SEWA-mediated application:

1. Real-time information of market rates.
2. Real-time daily wage rate.
3. Real-time weather with parameters useful for farmers (such as detailed rain forecast).
4. Instant messaging services that can connect them to cooperative leaders.
5. Group chats and the ability to create agriculture-related interest groups within the application.
6. Learning and sharing platforms on latest commercial and agroecological/non-commercial farming techniques.

Table 3: Data Needs Mapping

A comprehensive mapping of the aggregated data needs of women farmers took place across all FGDs. The data was then coded and classified into personal, non-personal and non-human data that cooperative members were willing to share.

Data Classification	Data Type	Consent to Share with Cooperative	Purpose of Use for Cooperative - Use Case	Comments
Personal Identifier	Name	Yes	For database creation and management	
	Age			
	Gender			
	Government Identity	Yes		
	Caste	Mixed	For internal census and for measuring targeted interventions	
Personal Data	Marital Status	Yes	For database creation and management	
	Household Status			
	Income Status, Personal and Household	No	To improve credit accessibility; For internal census and for measuring targeted interventions	Women across FGDs spoke about their fear of being classified into a higher income category, resulting in being removed from specific welfare schemes offered by the state.
	Agricultural Income	Mixed		
	Credit Information	No		

	Location	Mixed	For providing targeted services to each farmer; for internal census and measuring interventions; collected to offer aggregated data points in order to improve farm productivity	Women from specific villages fear misuse of location data for unfair profiling
	Enrollment in Govt Schemes	Yes	For database creation and management	
	Health Information			Current health status, support needed from SEWA, insurance status, etc.
Agricultural Data	Input Cost	Yes	For providing targeted services to each farmer; for internal census and measuring targeting interventions; collected to offer aggregated data points in order to improve farm productivity.	
	Input Materials Used			Includes information on: seeds, fertilisers, suppliers, varieties, etc.
	Input Time			
	Farming Techniques			
	Farming Troubleshooting Methods			With farmers from the same Mandli
	Soil Condition			To be facilitated by soil testing
	Income from Output	Mixed		Specific to income made from each agricultural season.
	Profit from Agriculture			
	Daily Wage Information			Some women are willing to share daily wage information, as long as they receive aggregated information on wages from the Cooperative

	Daily Market Rates	Yes	Aid in creating real-time aggregated market information.	
Agricultural Knowledge	Commercial Farming Techniques	Yes	To establish formal internal knowledge database.	With farmers from the same Mandli
	Non-Commercial Farming Techniques			
	Sharing from Capacity Building Sessions			With Cooperative as well as with farmers from the same Mandli

Table 4: Downstream Data Aggregation and Upstream Sharing

Upon determining women’s data needs, various case-uses of aggregated data were discussed within the consultation process, particularly those that benefit farmers through upstream sharing (cooperative to the farmer).

Data Classification	Data Type	Upstream Sharing (Cooperative to Farmer)			Purpose (Use-Case Examples)	Comments
		Aggregated at Potential Application (Village Level Information)	Aggregated at Mandli	Aggregated at Cooperative		
Aggregated Personal Data	Average Agricultural Income - Individual Level	Yes	No	No	Potential collective bargaining with panchayats for wage setting.	
	Average Agricultural Income - Village Level	No	Yes	No		
	Average Agricultural Income -Mandli Level	No	Yes	No		
	Average Agricultural Income - Cooperative Level	No	No	Yes		
	Aggregated Credit-related Information	Yes	Yes	Yes	Enable assessment of most popular/farmer-friendly banks.	Aggregated information on farmers who have accessed credit, banks used. Will not include information on repayment status, loan default, specific purposes of loans, etc.

Agricultural Data	Input Cost - Individual Average	Yes	No	No	Analysing individual input-spending trends over time.	Potential application can offer a farmer level profile with key information on input cost, output, profits, etc. Can also streamline information on soil testing, inputs bought through SEWA, etc.
	Input Cost - Village Average	No	Yes	No	Insights into village-level spending; can be used to assess interventions for making collective decisions at village-level.	
	Input Cost - Mandli and Cooperative Level	No	Yes	Yes	Insights into Mandli-level spending; can be used to assess interventions for making collective decisions at Mandli-level.	
	Soil Condition - Individual Trend	Yes	No	No	Available only to farmers over potential application for individual purpose, access to soil testing results.	
	Soil Condition - Village Level	No	Yes	No	Enable village-level action on soil condition; assess pesticide/fertiliser use, offer information that help farmers collectivise with institutional bodies (such as the panchayat).	

	Profit from Agriculture - Village and Mandli Level	No	Yes	No	Insights into village-level and Mandli-level profits; can be used to assess interventions for making collective decisions at village-level and Mandli-level.	Can be disaggregated by seed type, fertilisers used, etc.
	Profit from Agriculture - Cooperative Level	No	No	Yes	Offer information on Cooperative functioning.	
	Wage Information - Village Level	No	Yes	No	Wage information updated daily/weekly can offer collective bargaining advantages.	
Agricultural Knowledge	Commercial Farming Techniques	Yes	Yes	No	To establish a formal internal knowledge database that can aid hyper-local learning-sharing of information.	With farmers from the same Mandli
	Non-Commercial Farming Techniques	Yes	Yes	No		
	Sharing from Capacity Building Sessions	Yes	No	No		With Cooperative as well as with farmers from the same Mandli

Federated Techno-Institutional Models

The mapping of the farmers' data needs and aggregation present us with the ability to propose a data architecture model with underlying governance principles, determined by the socio-polity of the cooperative. To this end, the study proposes the usage of the term, techno-institutional as a prefix to the data lifecycle and architecture model; here, we see selective consolidation and incorporation of SEWA's existing data into the ontology of a techno-intervention, giving equal prominence to the economic principles of the cooperative, as well as those that guide the architecture of a data cooperative. Additionally, the data cooperative is an entity that is subsumed within the structure and existence of SEWA Cooperative Federation: by default, its governance, functioning and purpose, hence, will be driven by the existing, federated decision-making loci of the cooperative.

The “Vertical-Horizontal Federation” of the Data Cooperative: Foregrounding the Techno-institutional Model through a Critical Appraisal of Datafication

The inherent nature of datafication systems as explored are one of centralisation, emerging through the ontological and epistemological points of objectivism and positivism.

Afterall, datafication as a process is not just the transformation of analogue information into a data system through processes of collection and aggregation, but is the generation of different kinds of value from data. As Meijas & Couldry (2019) note, the process of quantification - presenting a new ontological stream - involves “abstraction via the process of turning the flow of social life and social meaning into streams of numbers”. This form of abstraction involves significant transformations to information through centralisations; through the caveats of big data systems within digital capitalism, the neoliberal logics and the faulty governance of such systems often lead to the formation of the “single source of truth” data paradigm, instituted to syphon data value, and ultimately relocate ownership to corporate-controlled digital intelligence (Gurumurthy et al., 2022b) Over time, data systems within digital and data capitalism only furthers the accumulation of value in the hands of digital monopolies, and in fact usurps the economic gains of smaller businesses, individuals and relevant data communities, where data value is captured “in the silos of rentier capitalism” (Gurumurthy & Chami, 2022a) and prevents the realisation of data's public and social value. The techno-institutional model must actively foreground the fact that it is not a base-up replication of such a system, but a radical alternative that embeds the imaginaries of a community-led resistance to the grasps of data capitalism. The problem is not data itself (Meijas & Couldry, 2019; Singh & Gurumurthy, 2021) but with how this data is collected and whom it is collected for; realising the redistributive capacities of data cooperatives and communities, hence, is directly a result of their

abilities to stand outside datafication, which entails critically viewing the centralising tendencies of data architectures and the adoption of a federated structure. The replication of such a digitalisation system through the incorporation of agricultural workers and their socio-economic foray from a top-down non-pluralistic techno-institutional model is at risk of magnifying the expropriation of data from the unorganised and informal workers of the Global South, who are often women workers from the bottom of the agricultural value chain.

In contrast to the mainstream digital systems within data capitalism, the skeleton of SEWA's economic and social relations is that of hyperlocalism and federation: the federation's "coming-together" and "holding-together" propositions are uniquely balanced, as noted in the section on SEWA's internal distribution of power and decision-making construction. Aptly, SEWA's data cooperative experiment has several consulting partners, each through their own guiding principles and analysis of data needs are constructing independently-moving parts of the data cooperative. The data cooperative project housed within the institution of SEWA hence is federated both horizontally and vertically; each Mandli, village and "unit of functioning" of SEWA's have varying data needs and gaps of accessibility. The assessment of these data needs by different research methodologies allows us to propose a variety of techno-institutional models that can collate to form a vertically-federated data cooperative. The intersection of these federated structures is a resistance to the imaginaries of a centralising, one-fit-for-all database (Delacroix & Lawrence, 2019), and a viable alternative to "single sources of truth" models.

The caveats of such a plurality, however, are the proliferation of differing and often divergent governance principles that determine the functioning of each techno-institutional model. Beyond the concern of interoperability that continues to remain a significant challenge to even digital intelligence systems controlled centrally, multiplicity in data systems can create "upstream challenges that can complicate downstream fairness" (Stoyanovich et al., 2017). In addition, these systems - as independent and as those that talk to each other - must ultimately be grounded through a data rights approach, enshrining data sovereignty (Calzada, 2021).

In addressing the former query, we propose the adoption of the *FAT by Design* approach (Stoyanovich et al., 2017). The principles of fairness, accountability and transparency amongst others within data science are often applied post-hoc (Boeckhout, 2018) - when the client or the *queriers* (Hardjono & Pentland, 2019) interact with the data system. the *FAT by Design* approach posits that datasets within any stewardship model must begin assessing for these properties as soon as datasets

with unknown biases and unclear provenances (Jethani & Fordyce, 2021) are collected, cleaned and integrated (Stoyanovich et al., 2017), and must persist through outside of the lifecycle of such a digitalised paradigm, with the potential to prevent the misapplication of wrongful databases which ultimately correspond to societal harms (Brown et al., 2018). The co-design process, hence, is an actioning of *FAT by Design*: a ground/base-up process of assessing data needs maintains historicity and data provenance, the process of “tracing and recording the origins of data and its movement between databases” (Buneman et al., 2000). The maintenance of data ancestry, historical origins and its movement across the database is not only a data architecture question, but a question of generating trustworthy data (Viglas, 2013); to that end, the *FAT by Design* approach to building a techno-institutional model aptly complements the consultative co-design process of “embedding trust” within the data layer.

Harmonising a Principles-Based Data Rights Perspective With a Base-Up Methodology: Enabling The Economic Redistribution of Data Value Among Unorganised Women Workers in Agriculture

In addition to *FAT by Design*, the protection of the rights of SEWA’s agricultural workers - who form the base of data value chain of every proposed techno-institutional model - also requires the confluence of a base-up research method with a top-down approach to establishing rights-based data governing principles. Propositioning a federated data cooperative for women agricultural workers - often from marginalised caste and class locations - requires the research process to approach institutionalising a data cooperative through a critical lens of political economy, and a rights-based approach to data generation, ownership and redistribution (Massey et al., 2023).

In this co-creation of an agricultural data cooperative, it becomes imperative to incorporate frameworks of data rights into its methodologies. SEWA both as a union and as a cooperative is pivotal to the social and solidarity economy and are harmonious with workers’ organisations with shared goals of promoting democracy and economic social justice (ILO, 2022).

The following proposed data governance principles have emerged from the convergence of existing legal frameworks that have seen their origins from decade-long, historical fights of trade unions (Public Services International, 2023) with the testimonies of women interviewed in the FGDs, who offer a rich contribution to the provenance of the data/s being collected across various collaborators. These transpiring principles are not only synergistic to the needs of SEWA as a federated, pluralistic cooperative, but can offer the path to building a data cooperative that protects against the transfer of data value to market-driven data structures.

1. Data Minimisation to Data Collectivism for Optimising Data Privacy

A general guiding principle of data minimalism for personal data ensures that data aggregators (as defined by the European Data Protection Supervisor) within the techno-institutional model should collect only data that is relevant to the purpose of the agricultural data cooperative. However, the aggregation of data into digital intelligence is key to distributing the economic gains of a data commons, and is pivotal in realising a regenerative appropriation of data (Gurumurthy et al., 2022c) that can enhance the cooperative's ability to improve the livelihoods of its workers, as well as optimisation of SEWA as an economic enterprise. Hence, the approach of data collectivism - in appropriate balance with principles of data minimalism - is a key framework to the co-design consultative process.

Data minimalism was often a non-negotiable precedent recognised by the women interviewed through the consultative process.

“Every locality needs to have its own version of the data cooperative. Not only should this be localised to the Mandli/land, but only data that can be of use to us ultimately must be asked for. Even the information provided to us must pertain to our livelihood and agriculture - we are not interested in providing all kinds of data so corporates can advertise to us.”

- AG⁵, during a focus group discussion in Vyara

2. Informed Consent and Optional Data Sharing

All data collected by the data cooperative must be done with informed consent of the farmer which is obtained through a consultative process by data officers/organisers who are trained by SEWA. Membership to cooperatives and SEWA cannot entail automatic consent to participation in the data cooperative. If certain services are to be completely streamlined through the data cooperative in the future, the benefits and the consequences of data sharing and entry of data into the cooperative's data life-cycle must be effectively explained to the farmers/worker-members.

The practice of informed consent, particularly through a trusted member of SEWA materialised as a key conduit to women's participation in the data cooperative. In Kheda, where the Mandli is currently in the process of re-forming after its dissolution during COVID-19, the comprehensive consultative process consisting of two federation members and two researchers from IT for Change proved to be a key trust-forming element in conveying the benefits that can emerge from the participation in a data cooperative.

⁵ Acronyms used in the research paper are fictionalised in order to protect the identity of the FGD participants.

“We generally are very wary of sharing data... the government collects so much data from us and we are aware of scams that happen because of sharing data. However, if someone we trust from the Mandli asks us to download an app or give our data, we definitely will. We will trust them because they have helped us before, and are part of SEWA”.

- RT, during a focus group discussion in Kheda

3. Anonymisation and Pseudonymisation

As data from workers travel downstream, the principles of anonymisation and pseudonymisation becomes an essential precondition to ensure robust personal data protection (AEPD-EDPS, 2021). Not only does an individual's data enter a data architecture, but is also intertwined into specific social systems that both govern data usage as well as offer unequal forms of data visibility. Particularly so, the rural agrarian context for women agricultural workers is well-situated in the confluence of a caste-class political economy (Sharma, 1984); protections of personal data in any endeavour of agricultural digitalisation must integrate appropriate data practice standards in order to protect farmers from being profiled and targeted (ILO, 2022).

The FGDs revealed that women agricultural workers within SEWA face significant challenges in accessing resources, markets and government schemes because of their caste and class locations; women feared that even the sharing of their village location with a data system could negatively impact them. Pseudonymised data was the only kind of information women were willing to share beyond the level of the Mandli.

“Most of us are not comfortable sharing personal income information - because of several issues with being taken advantage of. There is a distrust in several villages, where people have lost money after Aadhar information has been shared. They would share information on what type of crops, cost of crops as long as it's not connected to our identity. Companies might use that information and we're scared our information will be revealed to all.”

- WA, a board member during a focus group discussion in Vyara

4. Transparency and Accountability

The techno-institutional model must instil principles of transparency and accountability, by empowering its users with a series of rights and privileges, as well as creating protocols of accountability at both the technological and the institutional layer. To this end, The Office of the United Nations High Commissioner for Human Rights's (OHCHR) human-rights based approach to

data (ILO, 2022) on transparency and accountability are being proposed in this paper as guiding principles:

Transparency: Data collectors of the project, namely, SEWA Cooperative Federation and its research partners, including all potential third-party data analysts should provide clear, openly accessible information about their operations, including research design and data collection methodology.

Accountability: Data collectors and aggregators identified through cooperative leaders, current and all subsequent research partners and third-party data analysts are accountable for upholding human rights in their operations.

Field insights on transparency and accountability began to unfold over the course of various FGDs, especially as the researchers began to engage in incorporating field notes and the processes of reflexivity to modify subsequent FGDs, a key part of the grounded theory approach. Within time, much of these principles were unveiled as researchers gained information on previously-held consultative processes; farmers offered critical insights into the subsumption of the processes of accountability and transparency to the larger goals of data collection.

“We’ve had many people interview us before, and we don’t get to know what happens to that data. Many times, these large surveys take over half an hour, and we don’t even get to see the results of it. We would like to be told what this research is for before you begin asking us questions, and we request you to share any data you’ve already collected”.

- Several farmers across two focus group discussions in Vyara.

5. Restrictions on Data Offboarding, Sale and Re-use of Data

At all points of the data life-cycle (collection, analysis, storage and offboarding), there must be restriction of sale of data. The ILO (2022) expands this principle to not just the sale of data, but also the deletion of data and transfer of data sets with associated inferences and profiles for commercial purposes. The scope of the techno-institutional model as well as this paper is limited in mapping out principles of non-aggregated and aggregated personal and non-personal data, but the rapid ethnography aims to map out the goals and intentions of all member types at SEWA and provide recommendations of data sharing.

One of the paramount goals of the consultative process was to determine to what extent farmers would like to monetise their data; against the context of growing public and private digitalisation systems in India⁶, there is potential for data cooperatives to explore monetising data collected

⁶ AgriStack, an initiative of the Indian state, is an upcoming government-backed data exchange that enables the aggregation of digitised land records with farmer profiles and other agricultural data.

(Manohar et al., 2020) in order to keep the cooperative sustainable. However, farmers as well as SEWA Federation members were rightfully wary of data re-use and sale of data.

“We are not interested in getting advertisements from private companies on any application or through the Mandli. We want to give data only to SEWA to ensure we have better access to daily market rates and information on agriculture.”

- TS, during a focus group discussion in Vyara

“SEWA does not have any plans to sell or monetise data collected through the data cooperative. In fact, we are not even sure of whether data will get aggregated across Mandlis- the goal has always been to remain hyperlocal with our data, and ensure farmers can gain the most out of the information.”

- MA, a senior staff member of the SEWA Cooperative Federation

6. Collective Worker Rights to Aggregated Data

The framing of aggregated data as a collective resource that also has ownership situated with the workers must be protected through ensuring the data cooperative recognises a collective workers' right to aggregated data.

- **Access to aggregated information:** Access to aggregated agricultural information is a primary purpose of the data cooperative, and must be enshrined as a right, where each farmer has the right to access aggregated personal and non-personal (agricultural, farm, credit, input, output, etc.) data.
- **Prevention of targeting based on individual farmer-profiling:** Internal balances and checks within the techno-institutional model through anonymisation and pseudonymisation must prevent the unfair targeting of individuals based on sensitive personal data, personal data and non-personal data. Use cases of “fair profiling”, i.e., solely for the purpose of executing beneficial interventions (such as expanding subsidisation of seeds to specific villages) must be communicated prior to the onboarding of farmers. Such use cases must be clearly delineated through future consultative processes. In case data is profiled for commercial purposes, it falls foul of agreed principles and will breach the terms within which the cooperative operates.

A core insight that emerged from the field explorations is the need to realise the collective ownership of data through the data cooperative. Farmers, Aagewans and Kaarobaris collectively agreed on the

need for the data cooperative to be an redistributive entity that first and foremost improves the economic conditions of the farmers who are part of SEWA.

“Everything we collect through this cooperative should be for the protection of the farmer and farmer land - it should be for the security, safety and protection of the farmer land. It has to take into account the diversity of who are water dependent, rain dependent, and recognise that there is no one stop solution. And with all the data collected, we must all enjoy the benefits from it, not just those who can easily access technology.”

- OP, a board member during a focus group discussion in Vyara

A Federated Techno-Institutional Model for SEWA

In addition to gathering information on types of data that benefit women farmers and their livelihoods from SEWA, the rapid ethnographies were also designed to facilitate the ideation of a base-up techno-institutional model. As already seen, data stewardship models are often constructed through a top-down consultation of downstream data actors, where the focus continues to be the enablement of aggregated data sharing, either for the purpose of monetization, or creating open data resources. As elucidated by the overarching data rights perspectives, the primary goal of harm minimisation and data collectivism take precedence in designing a techno-institutional model. The goal of a federated structure of architecture is to create a base that is knitted to secure the interest of the members, but can then engage multiple federated structures.

As noted, this model is proposed to be one among several that have emerged from other independent consultative processes; interoperability of data and how this techno-institutional structure “talks” to other data systems is a work in progress that is not within the scope of this research. Additionally, several proposed parts of this data architecture model are under development; the imagination of a farmer-facing application may be modified as more comprehensive information on data accessibility arises through future co-design processes. Additionally, the federated architecture proposed will vary depending on the data capacities of Mandlis as well as village; the model by its nature allows for a hyperlocal adoption of its underlying principles while accommodating necessary changes at the architecture level.

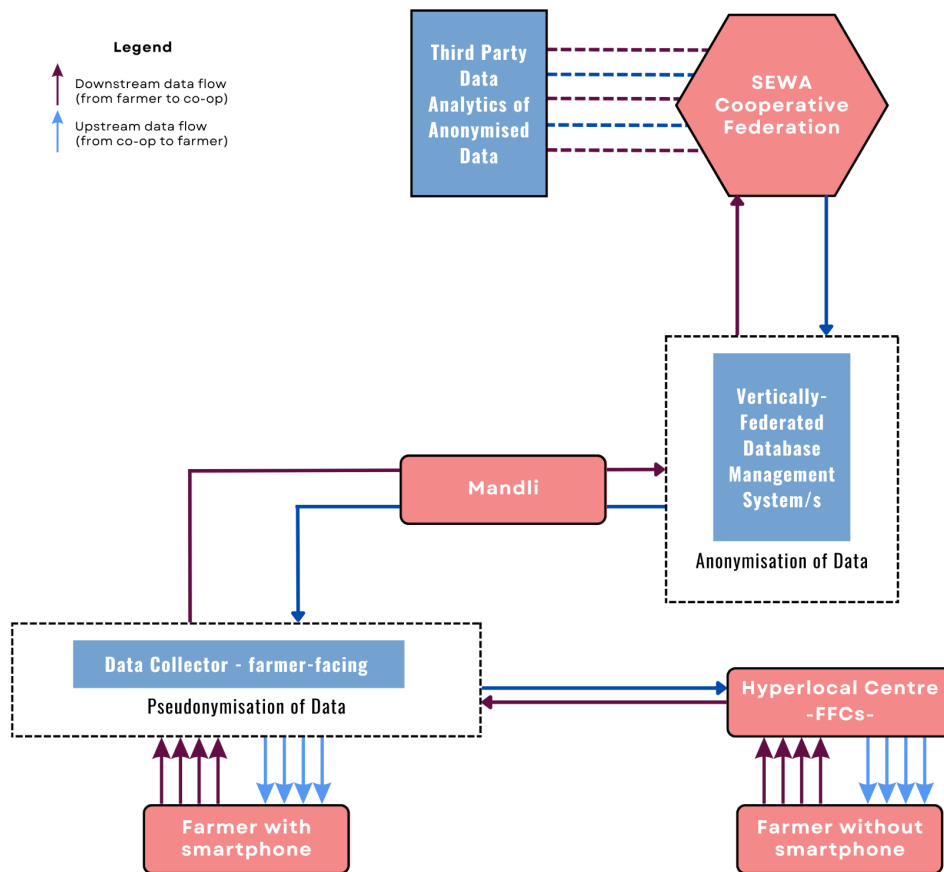


Image 1: A proposed internally-federated techno-institutional model for SEWA's data cooperative

The Data Generators: Farmers and Farmer Facilitation Centres

Women members of the SEWA are the primary data generators and owners. They enter the data life cycle of SEWA's data cooperative through a potential farmer-facing application/data collection process. The multiple categories of data as previously classified can be inputted into the database directly by the farmer. The baseline assessment conducted by IT for Change, in addition to the rapid ethnographies revealed that around half the farmers surveyed do not possess a personal smartphone but have access to an internet-enabled smartphone at the household level; over 25% of households also do not own a smartphone at all, effectively eliminating them from directly interfacing with the application.

As a systemic problem, the gender-rural digital divide has been framed as an active barrier to the functioning of the data cooperative. The proposed mitigation of this barrier comes from incorporating manual data practices and leadership structures of the cooperative into the techno-institutional model by embedding a physical-digital property to the techno-design; to this end, hyperlocal centres - also known as Farmer Facilitation Centres (FFCs) - already provide agri-informational services to farmers, and are often the one-stop-shop for accessing inputs provided

by SEWA, along with aiding capacity building. The FFCs hence, can be the points through which farmers who do not possess adequate digital literacy onboard onto the data collective, with the aid of Aagewans.

Pseudonymisation of Data

In interviews with farmers and multiple SEWA federation leaders, both the need for data localisation and data minimisation strongly emerged, with several farmers wary of sharing key personal identifiers (i.e. sensitive personal data as defined by India's proposed Personal Data Protection Bill, 2022⁷ as well as the Information Technology Rules, 2011⁸) beyond the federated level of the hyperlocal centres. In fact, women across multiple rounds of consultations were keen to see both data pools - the storage of data from multiple people - as well as data flow management emerge at the hyperlocal centre level.

In order to protect farmers from targeted profiling, the backend architecture of the proposed application/data collection system must enable automatic and mandatory pseudonymisation of data. Pseudonymisation of data under the EU General Data Protection Bill (GDPR)⁹ refers to the "processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information provided... which is kept separately". Pseudonymised data not only addresses overall personal data protection concerns, but specific requirements put forth by the data generators and owners in the institutional model. In one of the villages under the Megha Mandli, women were aware of the fact that service providers - both within and outside of SEWA - are often apprehensive of offering input services to their village as it largely consisted of an Adivasi population, in addition to being more remote. Hence, pseudonymisation of village names, for example, can be vital for the process of trust to be established and re-established within the data cooperative.

It is important to address that despite recognising the need for hyperlocal centres to also be the data stock and flow centres of all non-anonymised data, constraints in the form of infrastructural capacities at the level of each FFC has resulted in the recommendation of a model where Mandlis, and not FFCs receive pseudonymised data. With follow-up consultation processes with data architecture experts, the capacity to build a techno-institutional framework that is further federated to improve hyperlocal centres' abilities to be both "data pool and flow centres" can be evaluated. Additionally, the most effective means of pseudonymisation must be determined for this specific

⁷ The text of the proposed Personal Protection Bill, 2022 can be found [here](#).

⁸ The Information Technology Rules, 2011 can be found [here](#).

⁹ The definitions used in the GDPR can be found [here](#).

use-case scenario; as the data cooperative is self-funded, it must balance data protection with financial viability, foreseeing potential infrastructural limitations prior to application design.

As pseudonymised data moves downstream - the flow of data from upstream actors, i.e., the generators to the aggregators/processors - it enters a series of databases that are federated, both within this proposed techno-institutional model, as well as other institutional models.

Data Provenance is Key to Dispelling “Single Source of Truth” Architectures

The techno-institutional design must explicitly deviate from a “single source of truth” approach, where data is often proposed to be aggregated within a singular application database. As multiple research partners collect, aggregate, annotate and clean data, the data is stored within various systems of references and ontologies. In addition, farmers themselves receive data of many aggregations; as our data needs assessment reveals, different federated levels aggregate data in a unique manner, with insights from a FFC/Mandli level that require a specific relationship of trust with the data (as this data may directly inform the farmer’s decision-making process) versus cross-Mandli data that offer farmers a glimpse into the functioning of SEWA as a federation.

Data provenance - metadata that provides readers of data with information of the origin of data, i.e., data lineage - hence, takes the foreground, and a key step in embedding trust within a federated techno-institutional model. Provenance annotations (Stoyanovich et al., 2017) - where data’s origins are captured at every stage of data collection and aggregation - must then extend to every stage of the data cycle, as measure of *FAT by Design*. In the context of the data cooperative, provenance annotations not only allow the system to self-account for quality and robustness of data, but enables farmers to understand the hierarchy - and hence, mediate trust and the assessment of usefulness of the data they receive from the data cooperative.

Leveraging Mandlis as Data Intermediaries

Mandlis in SEWA are key federated units that oversee a series of hyperlocal centres and the villages that come under their jurisdiction, often making key decisions on factors such as subsidised input distribution and allocation, maintenance of input databases and facilitation of board member meetings. Within the proposed techno-institutional framework, Mandlis’ role as intermediaries between the data that is aggregated across Mandlis - at the level of SEWA Cooperative Federation’s research teams - and the Data Generators can be leveraged to ensure Mandlis also play the role of data intermediaries (Massey, 2022).

As a data intermediary, the application database management responsibilities lie with the Mandli. The following are proposed primary data and data management responsibilities of Mandlis:

1. Receive unaggregated pseudonymised data from multiple villages and hyperlocal centres.
2. Aggregate, clean, process, analyse and redistribute upstream the received data in order to enable farmers to gain more farm-level insights. Some use case scenarios for this include a trend analysis of harvest data for a new seed variety, changes in agricultural income as a function of improved access to fertilisers and other input services, etc.
3. Conduct Mandli-level needs and trends analysis of data received from all villages in order to create a PIN code-level aggregated profile of villages. Some use-case scenarios include evaluation of interventions (such as introductions of new subsidies on overall profits as measured at the village level), identifying barriers to productivity enhancement based on village-level performance, assessing soil conditions of specific areas/villages across time, etc.
4. Facilitate cross-village and cross-FFC comparison of agricultural data.
5. Anonymise data received from upstream data generators through the above-mentioned federated database management systems prior to the movement of aggregated data to the SEWA Cooperative Federation.

The techno-institutional framework recommends that the data processing capacities of all Mandlis within SEWA are enhanced, in addition to improving data protection mechanisms both at the technology layer as well as the human-intervention layer. The Mandlis in this model hold onto key responsibilities that ensure both the personal data of farmers are protected, as well act as conduits that execute the upstream distribution of data insights and hence, generate redistributive revenue that emerges from collectivisation of data. To that end, Mandlis are key to the accountability and fail-safe mechanisms of the data cooperative.

SEWA Cooperative Federation's Role as Data Processors

The last downstream unit in this federated model is the SEWA Cooperative Federation, acting as an Enterprise Support System through various internal teams. In the current structure of SEWA, several key cooperative-level decisions are often made at this level; data-driven insights, hence, can be of immense use in ensuring that the cooperative is self-sustainable, and assessing cooperative policies that can enhance women's economic gains.

At the level of each Mandli, data from the vertically-federated databases are anonymised to ensure personal data indicators are removed prior to the movement of data to the Federation level research team. Anonymisation was a repeated theme that emerged through the series of rapid ethnographies and consultations with upstream actors and data generators, and the process also federates and distributes data processing demand across two levels - the Mandli and the Federation, as the requirement of granular, de-anonymised data reduces as we move downstream.

On that end, the following are proposed primary and data and database management responsibilities of SEWA's research team:

1. Receive aggregated, anonymised data - either at the level of the Mandli or at the level of the village
2. Aggregate data across Mandlis to produce high-level insights on SEWA Cooperative Federation's agricultural workers.
3. Convert data from database management software/s to portable, large-scale datasets that can be used for both internal and external research. A use case scenario that can be elucidated here is the final conversion of raw, anonymised data of a potential SEWA agricultural census into multiple, analysable formats that can help the Federation in implementing changes to developmental interventions and internal directives.
4. Optionally share anonymised and cleaned datasets with third-party analytics partners if and when internal data capacities are not sufficient to produce actionable insights from datasets.
5. Implement upstream distribution of data in two forms: (a) as data insights in the form of datasets that Mandlis can analyse, and (b) as reports to all upstream parties (intermediaries as well as generators)

A key point of examination for the above techno-institutional model is to determine the type and extent of anonymisation measures that must be instituted at the level of the Mandli prior to the movement of data downstream. Beyond a determination of anonymisation techniques (Murthy, 2019) (generalisation, suppression or masking) that balance the needs for data minimisation as well as data collectivism, a second consultation at the level of Mandlis must also be completed in order to ensure the redistributive capacities of data upstream are not impinged by blanket anonymisation.

Conclusions and Challenges

The data cooperative experiment of SEWA Cooperative Federation and IT for Change is a radical proposition that aims to equip unorganised agricultural workers with the economic gains of aggregated data. Through our rapid ethnographic approach, the independently-moving parts of a federated cooperative were mapped, informing the need for an internally-federated data architecture system. Across eight consultative sessions, the distinct intersections between gendered access to data infrastructure and the experiences of caste-class marginalised women in an agrarian political economy became evident. This enabled a comprehensive mapping of the data requirements of women farmers within SEWA.

The success of the execution of this federated techno-institutional model is, however, not without significant challenges. A significant technological drawback of institutionalising multiple federated models brings in core challenges of interoperability; such queries can only be addressed through improving collaborations across various partners. The proposed model aims to infuse the federated spirit by ensuring personal data only reaches and is processed/pseudonymised at the level of Mandlis; however, data infrastructures of Megha and Kheda Mandli require vast transformations in order to circumvent more centralised data processing, which may present itself as a violation to the guiding principles of the data cooperative.

Data cooperatives face significant legal challenges as well: several countries do not have legal frameworks to legislate independent, cooperative data entities. Extensive research on the kinds of legal-institutional support needed by data cooperatives from the state are vital to ensuring data cooperatives and by extension, data trusts remain sustainable and scalable.

By foregrounding and harmonising (1) *FAT by Design* and a (2) principles-based data rights approach, we propose a techno-institutional model for SEWA's Data Cooperative, that is federated horizontally (the presence of multiple aggregators across the data lifecycle) and vertically (the presence of various data architecture models, as brought in by various consultative partners). Our proposed federated data cooperative aims to counter the centralising tendencies of datafication. It brings together existing data rights-based legal frameworks resulting from historical trade union struggles and the principles of collectivism embodied by SEWA's cooperative mode of production. The objective is to achieve a regenerative appropriation of data that enhances the cooperative's capacity to improve the livelihoods of agricultural workers through equitable distribution via data collectivism.

Bibliography

Bhattacharya, Tithi, (editor.) & Vogel, Lise, (writer of foreword.) & JSTOR (Organization) (2017). *Social reproduction theory : remapping class, recentering oppression*. Pluto Press, London

Boeckhout, M., Zielhuis, G. A., & Bredenoord, A. L. (2018). The FAIR guiding principles for data stewardship: fair enough?. *European journal of human genetics*, 26(7), 931-936.

Brown, A. W., Kaiser, K. A., & Allison, D. B. (2018). Issues with data and analyses: Errors, underlying themes, and potential solutions. *Proceedings of the National Academy of Sciences*, 115(11), 2563-2570.

Buneman, P., Khanna, S., & Tan, W. C. (2000). Data provenance: Some basic issues. In *FST TCS 2000: Foundations of Software Technology and Theoretical Computer Science: 20th Conference New Delhi, India, December 13–15, 2000 Proceedings 20* (pp. 87-93). Springer Berlin Heidelberg.

Calzada, I. (2021). Data co-operatives through data sovereignty. *Smart Cities*, 4(3), 1158-1172.

Chakraborty, S. (2021). (rep.). *Women in the Indian Informal Economy*. India Spend. Retrieved June 2023, from https://www.indiaspend.com/uploads/2021/03/26/file_upload-446784.pdf.

Delacroix, S., & Lawrence, N. D. (2019). Bottom-up data Trusts: disturbing the 'one size fits all' approach to data governance. *International data privacy law*, 9(4), 236-252.

ETC Group. (2022). Food Barons 2022: Crisis profiteering, digitalization and shifting power. *Springfield: ETC Group*. Available online at: <https://etcgroup.org/content/foodbarons-2022> (accessed June 15, 2023).

European Data Protection Supervisor. (n.d.). *European Data Protection Supervisor: Glossary*.

European Data Protection Supervisor.

https://edps.europa.eu/data-protection/data-protection/glossary/d_en#:~:text=Data%20minimizatio%20necessary%20to%20fulfil%20that%20purpose.

European Data Protection Supervisor. (2021). (rep.). *AEPD-EDPS joint paper on 10 misunderstandings related to anonymisation*. Retrieved June 2023, from https://edps.europa.eu/data-protection/our-work/publications/papers/aepd-edps-joint-paper-10-misunderstandings-related_en.

Fernandez, R., Adriaans, I., Klinge, T. J., & Hendrikse, R. (2020). (rep.). *Engineering digital monopolies: The financialisation of Big Tech*. SOMO. Retrieved June 2023, from <https://www.somo.nl/the-financialisation-of-big-tech/>.

Fraser, A. (2019). Land grab/data grab: precision agriculture and its new horizons. *The Journal of Peasant Studies*, 46(5), 893-912.

Gurumurthy, A., Chami, N., & Kumar, R. (2022b). (rep.). *Recasting Land Tenure Rights in the Data Epoch: Insights from a Country Case Study of India*. IT for Change. Retrieved June 2022, from https://itforchange.net/sites/default/files/2224/ITFC_Recasting%20Land%20Tenure%20Rights%20in%20the%20Data%20Epoch.pdf.

Gurumurthy, A., & Chami, N. (2022a). Governing the Resource of Data: To What End and for Whom? Conceptual Building Blocks of a Semi-Commons Approach. *Social Science Research Network*.

<http://dx.doi.org/10.2139/ssrn.4037323>

Gurumurthy, A., Chami, N., Chatterjee, S., & Shah, S. (2022c). (rep.). *Workers' Data Rights in the Platformized Workplace - A New Frontier for the Labor Agenda*. IT for Change. Retrieved June 2023, from

<https://itforchange.net/workers%E2%80%99-data-rights-platformized-workplace-a-new-frontier-for-labor-agenda>.

International Labour Organisation. (2022). (issue brief). *Improving Workers' Data Rights*. Retrieved June 2023, from https://www.ilo.org/actrav/pubs/WCMS_862219/lang--en/index.htm.

International Labour Organization. (2022). (rep.). *Resolution concerning decent work and the social and solidarity economy*. Retrieved June 2023, from

https://www.ilo.org/ilc/ILCSessions/110/reports/texts-adopted/WCMS_848633/lang--en/index.htm.

Jethani, S., & Fordyce, R. (2021). Darkness, Datafication, and Provenance as an Illuminating Methodology. *M/C Journal*.

Keller, J. (2021, June 25). What are 'bottom-up' data institutions and how do they empower people? [web log]. Retrieved June 2023, from

<https://theodi.org/article/what-are-bottom-up-data-institutions-and-how-do-they-empower-people/>

Kumar, R., Ganapathy, A., & Shah, S. (2022). (rep.). *SEWA Cooperative Federation Baseline Survey Report*. IT for Change. Retrieved June 2023, from

<https://itforchange.net/sewa-cooperative-federation-baseline-survey-report>.

Latonero, M., & Aneja, U. (2021). (rep.). *Co-designing Digital Interventions and Technology Projects with Civil Society*. World Economic Forum. Retrieved June 2023, from

<https://www.weforum.org/whitepapers/co-designing-digital-interventions-and-technology-projects-with-civil-society/>.

Manohar, S., Ramesh, A., & Kapoor, A. (2020). (rep.). *Understanding data stewardship: taxonomy and use cases*. Aapti Institute. Retrieved June 2023, from

<https://thedataeconomylab.com/wp-content/uploads/2020/06/Understanding-Data-Stewardship-Aapti-Institute.pdf>.

Martin, L. T., Nelson, C., Yeung, D., Acosta, J. D., Qureshi, N., Blagg, T., & Chandra, A. (2022). The Issues of Interoperability and Data Connectedness for Public Health. *Big Data*, 10(S1), S19-S24.

Massey, J. (2022, August). Retrieved June 2023, from <https://www.theodi.org/article/what-are-data-intermediaries/>.

Massey, J., Morinier, S., & Parkes, E. (2023). (rep.). *Responsible Data Stewardship*. Open Data Institute. Retrieved June 2023, from <https://www.theodi.org/wp-content/uploads/2023/03/Responsible-data-stewardship.pdf>.

Mehta, S., Dawande, M., & Mu, L. (2022, February 1). The key to designing sustainable data cooperatives. Retrieved June 2023, from <https://www.weforum.org/agenda/2022/02/the-key-to-designing-sustainable-data-cooperatives/>.

Mejias, U. A., & Couldry, N. (2019). Datafication. *Internet Policy Review*, 8(4). <https://doi.org/10.14763/2019.4.1428>

Murthy, S., Bakar, A. A., Rahim, F. A., & Ramli, R. (2019, May). A comparative study of data anonymization techniques. In *2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)* (pp. 306-309). IEEE.

Ole Borgen, S. (2001). Identification as a trust-generating mechanism in cooperatives. *Annals of public and cooperative economics*, 72(2), 209-228.

Papadimitropoulos, E. (2021). Platform capitalism, platform cooperativism, and the commons. *Rethinking Marxism*, 33(2), 246-262.

Pattnaik, I., & Lahiri-Dutt, K. (2022). Do women like to farm? Evidence of growing burdens of farming on women in rural India. *The Journal of Peasant Studies*, 49(3), 629-651.

Pentland, A., & Hardjono, T. (2020). 2. Data Cooperatives. In *Building the New Economy* (0 ed.). <https://doi.org/10.21428/ba67f642.0499afe0>

Public Services International. (2023, April). *Data Rights and Data Protection*. Data rights and data protection - PSI - The global union federation of workers in public services. <https://publicservices.international/digital-bargaining-hub/5-data-rights-and-data-protection?id=131>

Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*, 6(1). <https://doi.org/10.1177/2053951718820549>

Sandoval, M. (2020). Entrepreneurial activism? Platform cooperativism between subversion and co-optation. *Critical Sociology*, 46(6), 801-817.

Scholz, T. (2016). Platform cooperativism. *Challenging the corporate sharing economy*. New York, NY: Rosa Luxemburg Foundation.

Sharma, M. (1984). *Caste, Class, and Gender: Women's Role in Agricultural Production in North India*. [Office of Women in International Development], Michigan State University.

Singh, P. J., & Gurumurthy, A. (2020). (issue brief). *Data Sharing Requires a Data Commons Framework Law*. IT for Change. Retrieved June 2023, from <https://itforchange.net/sites/default/files/add/Data%20Sharing%20Requires%20A%20Data%20Commons%20Framework%20Law%20DGN%20Policy%20Brief.pdf>.

Singh, P. J., & Gurumurthy, A. (2021). (working paper). *Economic Governance of Data: Balancing individualist-property approaches with a community rights framework*. IT for Change. Retrieved June 2023, from <https://itforchange.net/sites/default/files/1880/Economic-governance-of-data.pdf>.

Stoyanovich, J., Howe, B., Abiteboul, S., Miklau, G., Sahuguet, A., & Weikum, G. (2017, June). Fides: Towards a platform for responsible data science. In *Proceedings of the 29th International Conference on Scientific and Statistical Database Management* (pp. 1-6).

Viglas, S. D. (2013). Data Provenance and Trust. *Data Science Journal*, 12, GRDI58-GRDI64.

