How to Build a Data Cooperative

A Practitioner's Handbook

Aapti Institute | Data2X | Open Data Manchester

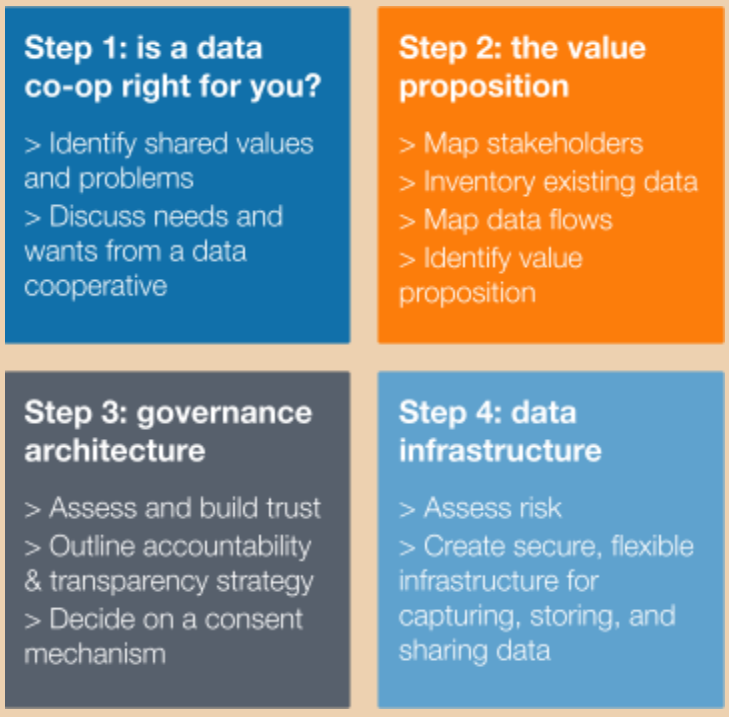
April 2024

Table of Contents

Executive Summary

Cooperatives—groups that pool their members' resources for a common goal—are a very successful model of economic organisation. Co-ops operate on seven key principles: 1) voluntary and open membership, 2) democratic control, 3) economic participation, 4) autonomy and independence, 5) provision of knowledge to other co-ops and the public, 6) cooperation with other co-ops, and 7) concern for the larger community. Over three million cooperatives exist across the world, employing 280 million workers, counting one billion people as members, and accounting for nearly 5% of global economic output. Worker, consumer, and financial cooperatives are the most common types of co-ops.

Newer to the scene is the data cooperative: a group of individuals or organisations that pool data for mutual benefit. Data co-ops serve many purposes, including protecting sensitive information, generating data-driven insights, and voluntarily sharing data for the social good and/or economic return to their members. This handbook is a hands-on guide for groups interested in forming a data cooperative. The process is laid out in four steps, all of which should be carried out in a participatory 'co-design' process open to all members of the group.



The first step is to evaluate whether the data cooperative model is a good fit for the needs of its members. The group should consider a variety of

factors, including shared values, needs, and wants among its members with respect to data. The data cooperative should represent a feasible approach to attain existing goals instead of being a 'solution in search of a problem,' as some tech-based innovations are. By the end of the first step, the group should have a good understanding of the strengths and weaknesses of data cooperatives, be able to decide whether a data co-op is appropriate, and, if so, plan co-design workshops to carry out the remaining three steps.

The second step is to discuss and agree upon the value proposition. This requires naming the data cooperative's purpose and reaching consensus on its fundamental values. The group should then assess and build basic data literacy skills to ensure that all members are able to participate. Next comes mapping of stakeholders—all the people and organisations with whom the co-op is likely to interact—and inventorying of all potential data resources to be managed by the cooperative. With stakeholder maps and data inventory in hand, the group is ready to map potential data flows, that is, with whom the data should be shared and for what purpose. The data flows exercise puts members in a position to specify the potential economic, social, and environmental value of pooling and sharing data.

After the value proposition is clarified, the third step is to design and establish an agreed-upon governance structure. Trust is the most important aspect of good governance, and the group must first assess the level of existing trust. Developing a plan for building trust over time may be necessary, especially if members hold sensitive data. Trust also depends on outlining a strategy that specifies how the co-op will embed accountability and transparency in its technological infrastructure, organisational design, and by-laws. Even existing co-ops with strong institutional frameworks will need to consider the particular challenges of *data* governance, including the development of consensus norms around privacy protection. One of the most important decisions the co-op will make has to do with the consent mechanism: the manner by which individuals give permission to share data, e.g., on a case-by-case basis versus less frequently. Consent preferences will likely differ based on the

type of data and end-user in question. Finally, the co-op should consider the governance implications of scaling up. Growing larger can place strain on a cooperative, especially with respect to maintaining trust, and having a plan in place will greatly facilitate a smooth evolution.

The fourth and final step in forming a data cooperative involves creating the data infrastructure. The first task is to conduct an assessment that identifies risks, outlines the potential consequences of data exploitation, and describes mitigation actions for each dataset. If the group intends to collect new data, it should also consider tools and methods for data collection that are secure, appropriate to data literacy levels, and lead to a smooth workflow. With a risk assessment and safe data collection strategy in hand, the cooperative can then decide whether on-premises or cloud-based data storage is more likely to meet their security, accessibility, and efficiency needs. The next task is to create safe, flexible ways of sharing data with end-user stakeholders. Questions of user interface and data access are important considerations. Finally, the cooperative should plan for steadily improving member participation in data collection and management. More inclusive processes build trust, data literacy, and confidence in the cooperative's value to members.

The success of data cooperatives ultimately hinges on the ability to join forces with others in the cooperative movement. This certainly includes other data cooperatives, but also traditional workers' cooperatives, consumer cooperatives, credit unions, and other allies. The future of data cooperatives is bright, but requires the creation of ever-wider circles of solidarity.

Acknowledgements

The primary authors—Astha Kapoor of Aapti Institute, Julian Tait of Open Data Manchester CIC, and Bapu Vaitla of Data2X—wish to thank our colleagues Shefali Girish, Sushmitha Viswanathan, Vinay Narayan, Megan Avery, Gratiana Fu, Neeraja Penumetcha, and Kelsey Ross for their substantial contributions to this work. We also wish to thank Kirsty Styles, Sam Milsom, members of the Data Cooperative Working Group, Krista Jones Baptista, and Lauren Pan. Thank you to the numerous experts who agreed to be interviewed for this publication, including Sara Mas Assens, Michael Dillhyon, Dazza Greenwood, Morshed Mannon, Trebor Scholz, and Janis Wong.

We thank the Rockefeller Foundation and Dalberg Catalyst for their generous support under the Uncommon Collaborations initiative, with particular thanks to Juliana Lopes Sauaia, Aaron Mihaly, Matt Freeman, and Robin Miller. We further thank the Brookings Institute and Rockefeller for their support of Room 9 of the 2022 '17 Rooms' initiative, with special thanks to Nathalia dos Santos, Alexandra Bracken, Sarah Geisenheimer, Zia Khan, and John McArthur, in addition to our Room 9 members Eliane Ubalijoro, Jeni Tennison, Kevin O'Neil, Kippy Joseph, Lucy Harris, Neema Iyer, Rabeh Ghadban, Revathi Kollegala, Rikin Gandhi, Salonie Muralidhara Hiriyur, Suneeta Krishnan, Sushant Kumar, and Varja Lipovsek.

List of boxes

List of figures

List of tables

Step 1. Determine if data co-ops are the right model for you.

1.1 A brief primer on data cooperatives.

Cooperatives are organisations that pool their members' assets for mutual benefit based on a common set of principles (Box A). Workers' cooperatives pool labour, while credit unions and consumer co-ops pool capital.

The emerging data cooperative model pools information, especially digital information. Such information includes data that is passively generated as we use digital technologies—for example, records of financial transactions, cell phone calls, and energy consumption statements. Other forms of data, such as health records and agricultural prices, must be actively collected.

Data has economic and social value. Some of this value is captured by

Box A: The cooperative principles

Cooperatives are a very successful form of economic organisation. Over three million cooperatives exist globally, employing 280 million workers, around one-tenth of the world's working population. Cooperatives count one billion people[1] as members,and hold assets of at least 20 trillion USD[2]. Revenue from cooperatives is close to 5% of global economic output.

The cooperative movement adheres to seven core principles[3]:

1. Voluntary and open membership to all people willing to actively participate.
2. Democratic control by members.
3. Economic participation by members: allocating resources for mutual benefit.
4. Autonomy and independence, partnering with others on an equal footing.
5. Education, training, and information for other cooperatives and the public.
6. Cooperation with other cooperatives—locally, nationally, and globally.
7. Concern for the community in which the co-op works.

companies and governments in the process of creating goods or providing services. However, much of the value of data goes unrealized because data is not accessed by potential users, is inefficiently collected or stored, or because its value has not been fully discovered.

Data cooperatives exist to unlock and distribute the value of data more equitably. The total global value of data is not precisely known, but it is likely highly concentrated in a few hands. For example, the early 2024 combined market capitalization of just five data-driven companies—Apple, Alphabet, Microsoft, Amazon, and Meta—was nearly eleven trillion USD. Data cooperatives allow the people who generate data to regain control over its value.

Data also has unrealised social value. Digital datasets can fill knowledge gaps and drive the creation of more equitable, effective public policy, especially for meeting the needs of marginalised populations[4]. Rigorously assembled datasets can also correct bias in the private sector; one example is the use of credit-scoring algorithms[5] to facilitate women's access to financial services.

Data cooperatives in particular offer individuals the ability to decide how entities like companies and governments access and use their data, or whether they access it at all. In the broadest sense, data is a form of representation. Controlling data about our own lives allows us to better represent ourselves in political and economic spaces. This potential is especially relevant for marginalised people across the world who may be invisible to, and sometimes deliberately excluded from, the social benefits to which they are entitled. Data cooperatives can help people become visible on their own terms.

Data cooperatives focus on different types of data, depending on their purpose and the resources available for collecting and managing data. An energy cooperative, such as Carbon Co-op[6], based in Manchester, United Kingdom, might collect data representing energy usage and efficiency. An agricultural cooperative, such as Megha Mandli[7] Cooperative in Gujarat, India, might collect and share credit history data to enable better access

to finance (see Box B at the end of this chapter). Other examples of data cooperative value propositions appear throughout this handbook.

This handbook walks organisations through the step-by-step process of building a data cooperative: identifying the value proposition, designing the governance system, and creating a data infrastructure. This first chapter deals with the first and most important step of all: determining whether the data cooperative model is a good fit for the needs of your organisation.

1.2 Evaluate whether a data co-op meets your needs.

The checklist below is offered as a tool to evaluate whether a data cooperative is appropriate for your organisation. A 'no' answer to any question does not necessarily mean that a data cooperative model is not a good choice, but rather may signal a key area on which to focus during the initial stages of planning the data cooperative.

- ☐ Does your group have a shared purpose and a set of common values known and acknowledged by all members?
- ☐ Do these values align with the seven cooperative principles listed in Box A?
- ☐ Does your group have strong existing bonds of trust, and is there a strong commitment to maintain and deepen this trust?
- ☐ Is there a clearly defined problem for which a data cooperative—i.e., pooling data for mutual benefit—is a possible solution?
- ☐ Do you have a good grasp of member needs and wants with respect to a data cooperative?
- ☐ Do you have adequate human and financial resources to dedicate to building and maintaining a data cooperative?
- ☐ Do your members possess basic digital literacy skills, or does your organisation have the resources to build these skills?

- ☐ Does your organisation know which types of data are currently being generated and stored?
- ☐ Do you have a clearly defined set of data users, either internal or external to your organisation?
- ☐ Do you currently have an effective governance system with high levels of member participation, especially with respect to consent over use of member assets?

Each question should be posed again several times within the co-design process to elicit feedback from members.

If a data cooperative does not meet the organisation's needs, other models for responsible and democratic data stewardship could be considered. Additional resources that describe these options are available at the Ada Lovelace Institute[8] and the European Commission's Joint Research Centre[9], among other places.

1.3 Plan the co-design process.

The co-design process for creating the data cooperative must be inclusive and reflect the interests of the organisation's members. Co-designing is based on the principle that members are experts in their own lived experience. Broadly, it refers to a process, a set of principles, and practical tools that help foster cooperation and trust among members. Co-designing is also aligned with the concept of data justice[10], as it increases the opportunity for people to be represented and treated fairly in decisions involving use of their data. When realised in its full potential, the co-design process is a means to not only decide on the details of a project, but also to discuss the broader efforts in which the project is embedded[11]—in the case of data cooperatives, the movement to shift power in data governance.

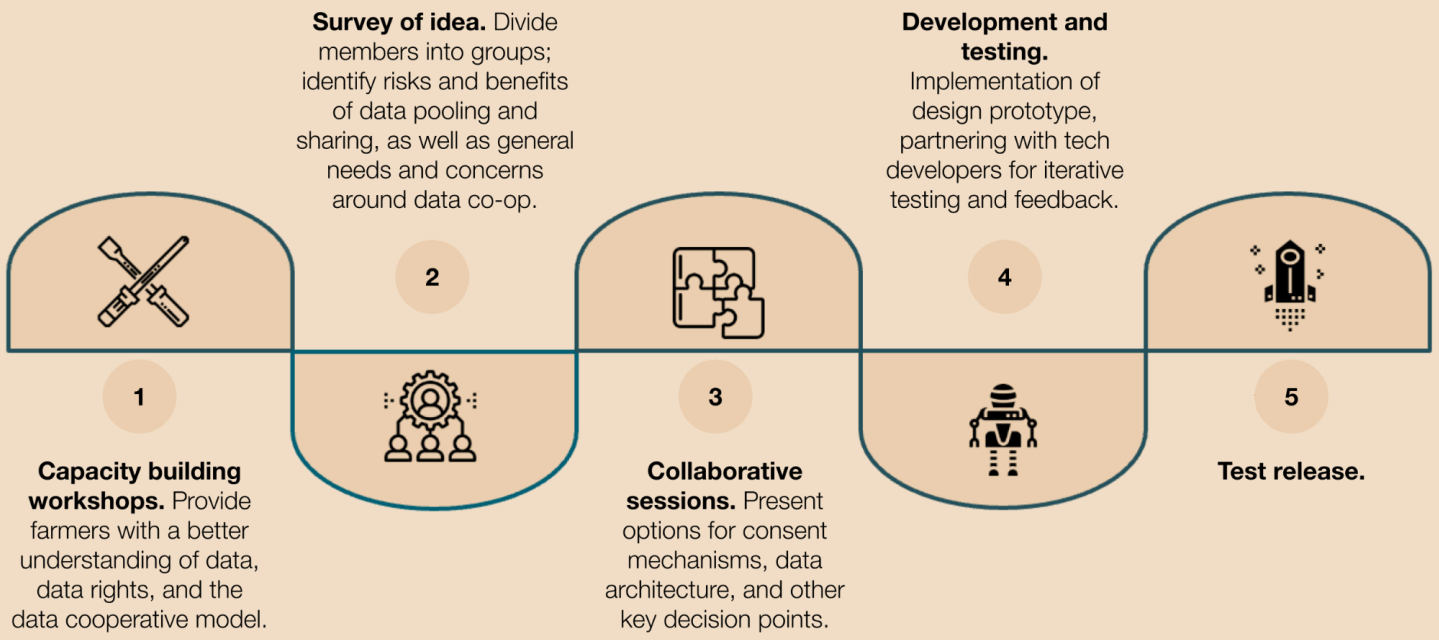
Practically, co-designing a data cooperative often entails a series of community workshops. Detailed toolkits for carrying out this co-design

process[12] are readily available. Here we outline a few general principles for planning and running these workshops:

- Be cognisant of logistical challenges faced by individuals and plan to meet them. The cost and time involved in transit, food, child care, accessibility, translation, and data literacy can be considerable. The organisation should have a plan for addressing these needs.
- Leverage community leadership. Including community leaders early in the process helps ensure that the co-design workshops are appropriate to the cultural, political, and economic context. Both formally structured meetings and informal events can help build relationships and generate ideas on how to create a non-intimidating space for collective judgement and deliberation.
- Engage participants in compelling, interactive ways. Encouraging storytelling and other familiar methods of expression can help create a collegial, friendly environment while offering individuals the opportunity to express themselves in diverse manners. Having a trained and experienced facilitator for the co-design sessions is helpful to keep participants engaged.
- Plan multiple sessions to iterate on and test ideas. Create a co-design process that proceeds at a pace comfortable for members, and offers space for reflection and revision. Remember that ideas will need to be tested and changed. Box B describes a co-design process carried out by the Megha Mandli cooperative in Gujarat, India.

Box B: The Megha Mandli co-design process

The Megha Mandli cooperative in Gujarat, India counts around one thousand indigenous women farmers as members. It is currently engaged in building a data cooperative layer that will pool the income and credit history data of its members to improve collective credit-worthiness vis-à-vis financial service providers. The diagram below illustrates the ongoing process for the co-design workshop and later implementation/testing.



Step 2. Identify the value proposition.

2.1 Name your purpose and affirm fundamental values.

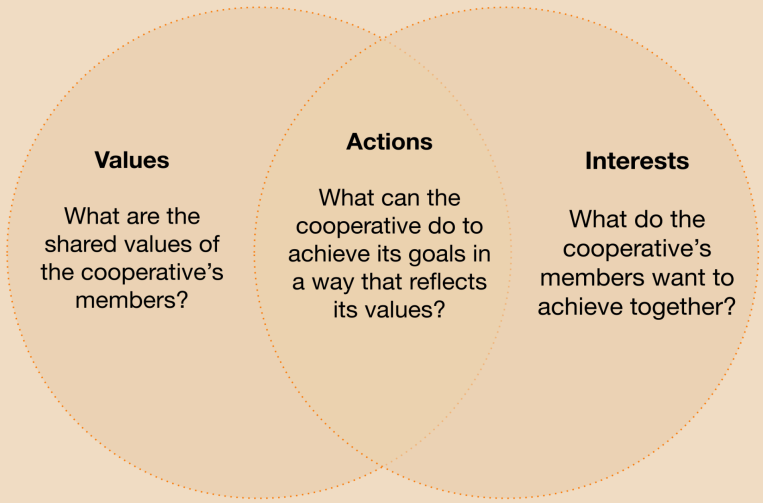
The first task in the co-design process is to clarify the problem(s) for which data cooperatives are a potential solution. The defined problems should relate closely to the core mission of the group. Be wary of deploying the data cooperative framework as a solution in search of a problem—an all-too common occurrence in tech innovation settings—which will result in 'identifying' problems that are peripheral to the organisation's mission. Make explicit why (or why not) the members of your group feel that a data cooperative is a good solution to the problem.

After naming the problem, it is useful to affirm the shared values that bind the group together. Revisiting these values can reveal whether a sufficient majority of the group's members feel that a data cooperative is not only a solution to the named problem, but a solution that reflects shared values. The values conversation should begin by naming the general values underpinning the organisation, and then proceed to engage specifically with values around data. Do members have privacy concerns about collecting and sharing data? Do members trust the organisation to safely manage their data? With which kinds of actors do members wish to have data sharing relationships—public sector, civil society, companies—and what are the ground rules for interactions in these relationships?

It is also important to discuss the distinction between values and interests. Values are fundamental principles that guide personal and organisational behaviour, whereas interests are the benefits that the person or organisation hopes to obtain. The intersection between values and interests helps define the group's course of action (Figure 1).

Many questions arise from considering the relationship between values and interests: what motivates members to join and stay involved in the organisation, and what would compel members to participate in a data

Figure 1. The intersection of values, interests, and actions.



cooperative? How do deeply held ethical values interact with economic interests? Differentiating between values and interests helps move the conversation past commonly held values and norms ('open data', for example) and into trickier, more fruitful conversations about trade-offs (what degree of openness fits with both privacy values and economic interests?).

The values conversation helps prepare the ground for decision-making in all the stages that follow. These later decisions will be regarded as transparent if they can be traced back to agreed-upon values. In addition, as the process moves forward, members may find new points of agreement and disagreement in values and interests. An early discussion about values helps create a space for such recurring individual and group reflection throughout the co-design process.

2.2 Assess and build data literacy.

Not all members of your organisation have the same familiarity with data and the use of digital devices. The potential of technology to dehumanise by exclusion must always be in view; discomfort with, and distrust of, emerging technologies deters engagement of members[13]. Building equity in data literacy is a long-term goal, but the co-design process can build a strong foundation. The first step is to ensure that all members have an adequate grasp of foundational concepts like data (Box C).

Box C: What is data?

When we think of data, we often think of computers and other digital technologies. But data exists in many forms and places—a written note, a photograph, or even a memorised fact all contain data. Most generally, data is any sequence of symbols that contains interpretable information.

Data consists of variables and values. Variables are symbols (most often expressed as words or letters) whose value is not fixed. For example, velocity (mph) is a variable and '6' is one possible value associated with that variable; and animal is a variable for which 'lion' is a possible value. Variables come in two forms: quantitative and qualitative.

Quantitative (also called 'numerical') data uses numbers to capture phenomena like cost and temperature. Quantitative data can be continuous, taking on any value in a range, or discrete, in which only certain values within a range are possible. Distance in metres is a continuous variable, while the number of people in a queue is a discrete variable.

Qualitative data uses linguistic or pictorial symbols to describe aspects of our lives that cannot be quantified. Qualitative data comes in three forms: 1) nominal variables which have no natural order among the values, such as gender, eye colour, and name; 2) ordinal variables, such as exam letter grades or product quality, which can be arranged on a scale; and 3) binary variables, which only allow two values such as 'yes' or 'no.'

It is also important to understand how data relates to information, knowledge, and wisdom. Data becomes information once it is interpreted; information becomes knowledge when it is deployed towards an objective; and knowledge becomes wisdom when it is used to make the 'right' choices. For example, digital temperature data may be stored as a series of numbers, perhaps zeroes and ones. It becomes information when the meaning of the pattern is interpreted as a value readily understood by human beings, e.g., 'the temperature is 42 degrees Celsius.' This information becomes knowledge when we contextualise it, e.g., 'physical exercise in temperatures above 40 degrees Celsius can lead to severe dehydration.' Knowledge becomes wisdom when it is connected to a normative goal, e.g., 'to avoid dehydration, wait to exercise until the temperature falls.'

In part, the co-design process is a means to create a level playing field for decision-making. Communities are more likely to engage[14] and design solutions if they understand how data and technology can be used for problem solving (Box D). Participatory engagement[15] engenders a sense of ownership as members themselves identify problems, generate, test, and refine ideas, and reveal inequities to be addressed. Inspiration can be taken from the success not only of cooperatives but also citizen science initiatives[16], many of whom have piloted new forms of horizontal decision-making. The overall goal is to design technology in a way that encourages, not inhibits, sustained participation.

Box D: Co-design of digital platforms

Residents of the Helsinki, Finland region were involved in the co-design of a smart city service[17] related to sustainable commuting. The project used a living lab model that included an iterative process which engaged users in preliminary planning to understand their needs, analyse the usability of proposed applications, and develop the application itself.

Abalobi[18] is another remarkable example. The South Africa-based project empowers fishers through co-created digital technologies and data analytics products. The process[19] helps create a sense of local ownership. Digital literacy proved to be an important metric of success; over time, fishers became more confident with their technological knowledge and skill.

A few key principles underlie the assessment and development of data literacy. First, the co-design process must continually monitor the inequities in data literacy that exist in any group. More data-literate individuals can dominate the cooperative. This undermines trust and engagement, and also lessens the value of the group's data assets: disproportionate representation of data-literate individuals introduces bias into datasets and leads to tensions about appropriate remuneration relative to data contributions. Conversations between all members about the balance between data privacy, member economic gain, and the public interest must be ongoing.

Second, in the short-term the organisation should strive to create user experiences matched to existing skills. The emerging field of community technology[20] explores the set of enabling conditions for meaningful broad-based engagement with digital technologies. At the most basic level, terminology and processes must be adapted to allow all members to understand what the data says and how the organisation could potentially use the data. For instance, using props and visual representations to facilitate co-design and utilising components that require minimal reading proficiency—e.g., using icons, checkboxes and other intuitive features—can be useful. A concerted effort to identify biases that occur in the process of technological development is also critical; GenderMag[21], for example, is a method that finds and fixes gender inclusivity bugs in software interfaces and workflows. The overarching principle in choosing tools is accessibility: all members of the cooperative should feel empowered to participate in modifying tools. This requires an understanding of local practices that can be gained only by relationship-building during the co-design process.

Third, intentional efforts to build data literacy over time are critical, including implementing a monitoring plan to track digital access and participation.The fifth cooperative principle (see Box A) focuses on education and training; cooperatives should seek to create an environment of continual learning. Data cooperatives in particular should empower people to make informed choices about how their data is used. This means building and maintaining confidence in the cooperative's systems of data storage, processing, and sharing. Such broad-based confidence also helps the most vulnerable members hold the cooperative accountable.

2.3 Map stakeholders.

With the cooperative's purpose and values defined, the group can begin to identify 'stakeholders'—people and organisations—with whom the cooperative is likely to interact.

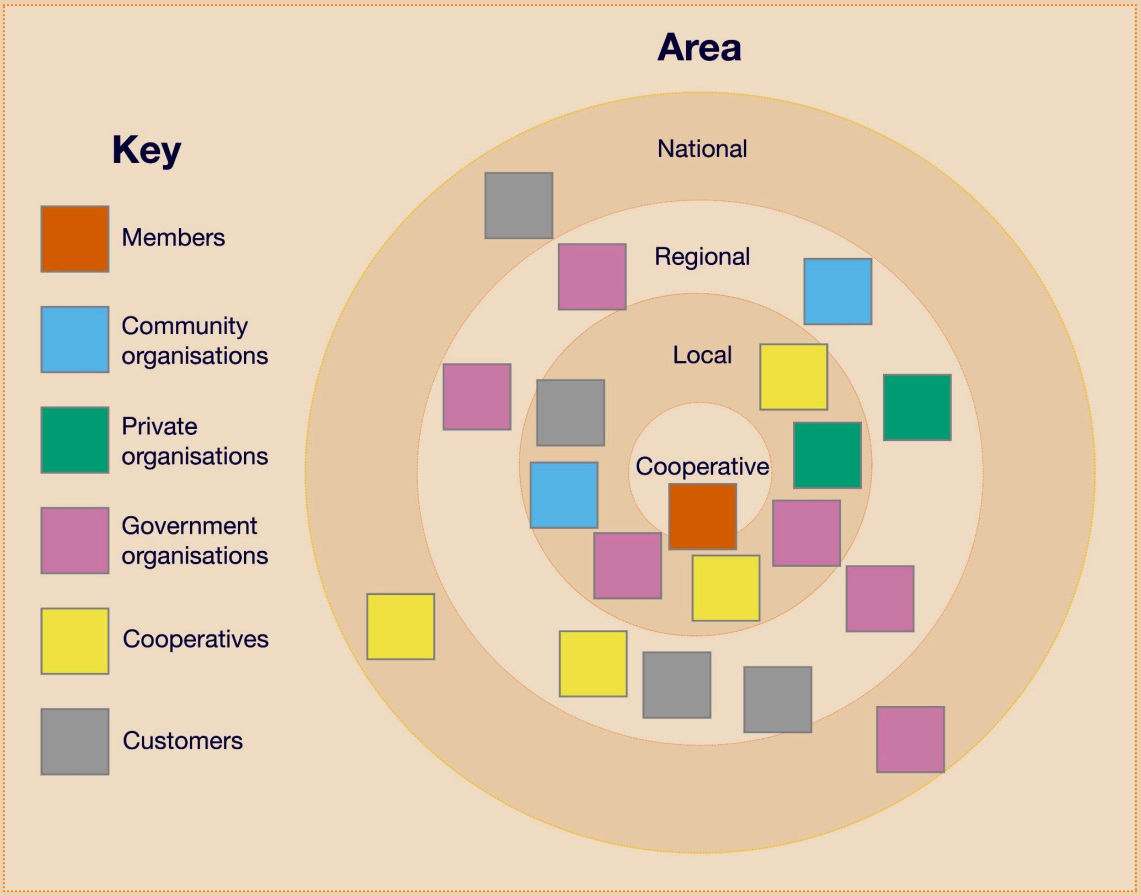
A useful starting point is to map anyone who may have a key relationship with the cooperative. Mapping can be undertaken in various ways, for example by brainstorming a list by category (Table 1) or geography (Figure 2). Online tools like Miro can help, although offline materials like sticky notes, paper, or a whiteboard will usually be sufficient.

The mapping work is best done by the cooperative's members, but could also include other trusted parties. It may be useful to map as many existing and potential stakeholders as possible first and pare down the list later. The paring down process will again be informed by the problems and solutions most relevant to the data cooperative, as well as the stakeholders associate with those solutions.

Table 1. Example of stakeholder mapping by simple list.

STAKEHOLDER CATEGORIES					
Cooperative members	**Suppliers**	**Customers**	**Financial service providers**	**Other cooperatives**	**Government agencies**
Member A	Raw materials supplier	Customer	Bank	National cooperative	Tax agency
Member B	Equipment supplier	Customer reseller	Credit union	Other cooperative 1	Regulator
Member C	Energy provider	Third-party customer	Local lender	Other cooperative 2	Sectoral support org

Figure 2. Example of stakeholder mapping by geography.



Note that the stakeholder mapping will raise an important question about the membership of the data cooperative: the single-stakeholder versus the multi-stakeholder approach. Single-stakeholder data cooperatives are more like traditional cooperatives: they have a flat structure with no recognisable hierarchy, and the members are those that generate the data, i.e., the data subjects. Multi-stakeholder cooperatives, on the other hand, can include as voting members both data subjects and data users, for example the organisations or government agencies seeking to utilise member data. The multi-stakeholder model may be useful in some circumstances, as it provides a means to balance the interests of data subjects and users. However, this handbook is oriented towards the single-stakeholder model, wherein membership is made of data subjects with equal voting rights.

2.4 Inventory data.

Next, take stock of the cooperative's data—not only the data to which it already has access, but also data that *could* be collected or created. Resources[22] are available to help you carry out a data inventory; this section summarises key aspects of the process.

The first step in undertaking a data inventory is defining its purpose and scope. The purpose informs the scope of data that the organisation will focus on for cooperative management, and should be rooted in seeking solutions to the problem(s) identified earlier in the workshop.

Data held by the organisation likely exists in many different formats—for example, a spreadsheet stored on a hard drive, a photograph on a phone, or a note written on a piece of paper. The initial stage of the data inventory process should itemise all forms of verbal, written, and digital data that relate to the identified purpose; miscellaneous data assets will be inventoried later.

Once datasets have been assembled, the following metadata should be generated for each dataset:

- Identifying information, e.g., a unique identification code;
- The name, format, and description of the dataset, as well as associated keywords;
- The time and place the dataset was created, frequency of updating, and date of latest update;
- Purpose of the dataset;
- The creator and current manager of the dataset;
- Location of dataset storage;
- Rights and restrictions that apply to the dataset.

This metadata should be stored in a separate associated file. In addition to metadata, a codebook should be created for each dataset, although this

can happen at a later stage of the co-design process. The codebook should contain a description of each variable in the dataset and how values are to be interpreted.

Once all relevant datasets have been itemised and metadata created, the organisation should review the quality of each dataset. High-quality data is easier to use and creates more value. Data quality has many aspects, including:

- Completeness: is the dataset complete enough to be useful, with respect to the topic(s) it covers?
- Accuracy: does the data reflect precisely what it seeks to represent, or does it contain frequent errors?
- Consistency: does the data record the same variable in the same way every time (e.g., remains quantitative within a given range)?
- Uniqueness: does the data contain duplicates or overlaps?

Note that this is an incomplete list. The workshop participants may wish to identify other features relevant to data quality.

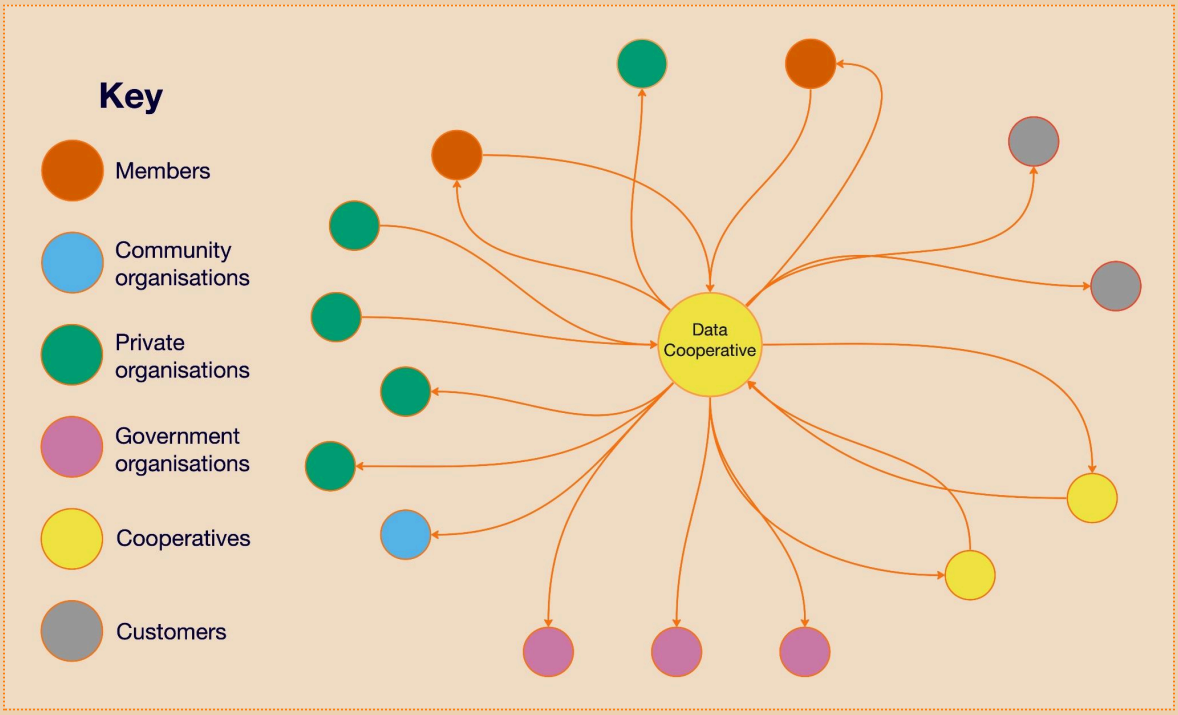
Ethical questions around data collection, storage, and sharing may arise during the data inventory. The fundamental tension is this: if data is locked behind a too-stringent control mechanism, it becomes difficult to use, but data in the wrong hands could cause harm. It's important to note members' views about the right balance between security and openness to inform a more detailed discussion of data risks and consent later in the co-design process.

2.5 Map potential data flows.

Once the key stakeholders are identified and the data inventory completed, we can create a 'data ecosystem map'. This map draws all existing and potential flows of data and information between the data cooperative and its stakeholders, as well as the 'value flows' that may be associated with each relationship (Figure 3). For now, consider 'value' to

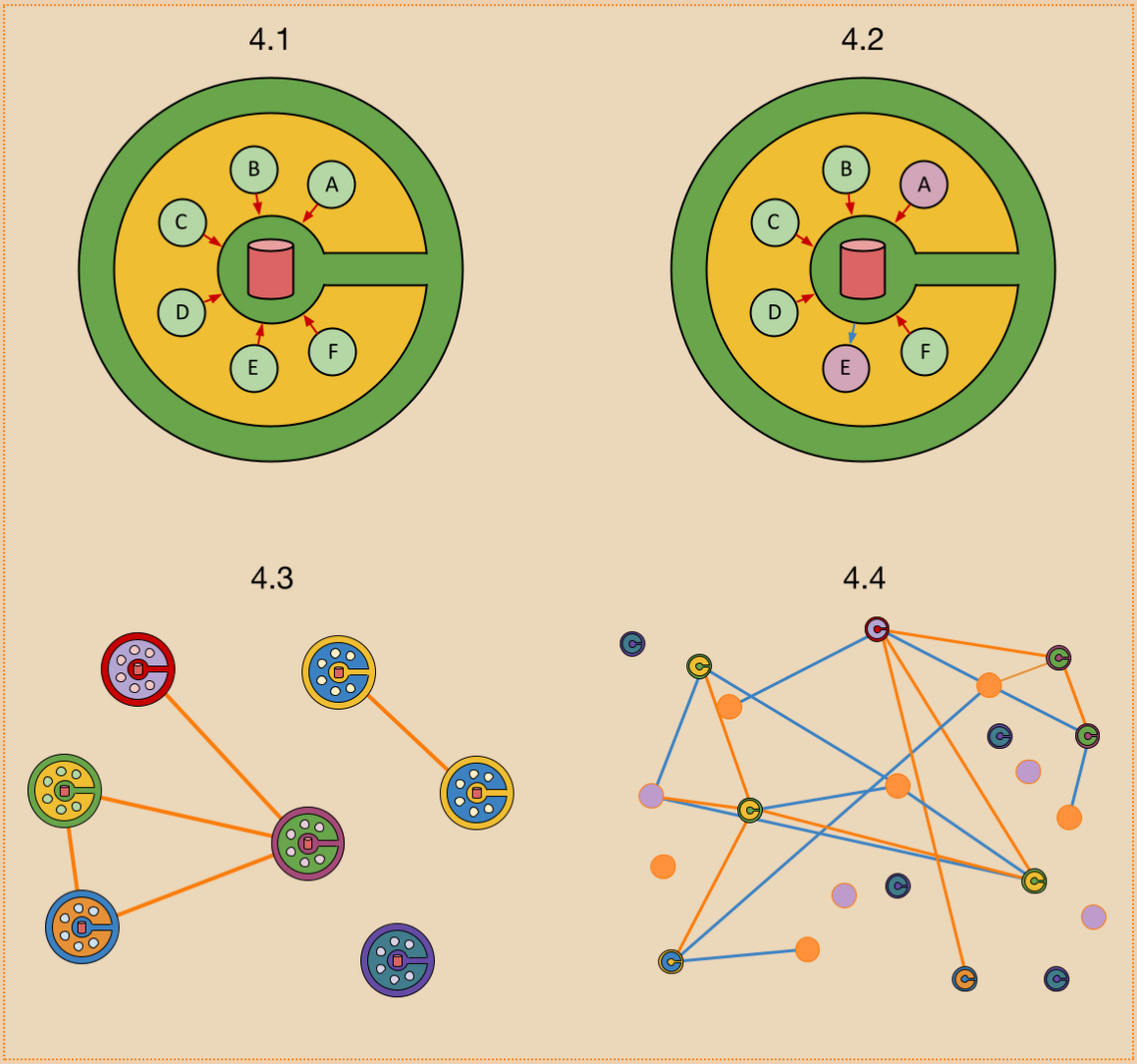
be a general term; in Section 2.6, we discuss how the data cooperative may choose to define value and negotiate trade-offs between different kinds of value—for example, economic value versus social value. The key exercise for now is to identify flows of data that offer any kind of value, as well as the direction in which the value travels (i.e., who's offering value and who's receiving value). When drawing the map, it is a good idea to put the data cooperative in the centre and stakeholders around it.

Figure 3. Mapping data flows between stakeholders.



Once a draft data ecosystem map with value flows is completed, the cooperative can begin to engage with the question of how it wishes to engage in data sharing.

Figure 4. Types of data flows. 4.1: Member-to-data-cooperative flow. The data cooperative is depicted by the database icon, and the cooperative members are labelled A-F. Data flows from individual members to the cooperative's database. 4.2: Intra-cooperative flow. Data is transferred from member-to-member via the cooperative's central database. 4.3: Federated data flow. Cooperatives with shared values and interests exchange data. 4.4: Third-party flow. Data is shared between cooperatives (circles with a 'C') and non-cooperative stakeholders (filled in circles).



Most generally, there are four types of data flows that could be handled by the data cooperative:

1. **Member-to-data-cooperative flow** (Figure 4.1) is the simplest type. Members (labelled A-F in the figure) share their data with the data cooperative (depicted by the database in the figure) for internal

use, especially to inform the development and delivery of the cooperative's services. This type of flow will likely already exist within established cooperatives, even if they have not yet created a formal data cooperative layer.

2. **Intra-cooperative flow** (Figure 4.2) occurs when data is shared between individual cooperative members, with sharing facilitated by the cooperative, possibly through a central database. For example, agricultural price information from members could be collected and shared to help each individual find the best opportunities for input purchase and output sales.

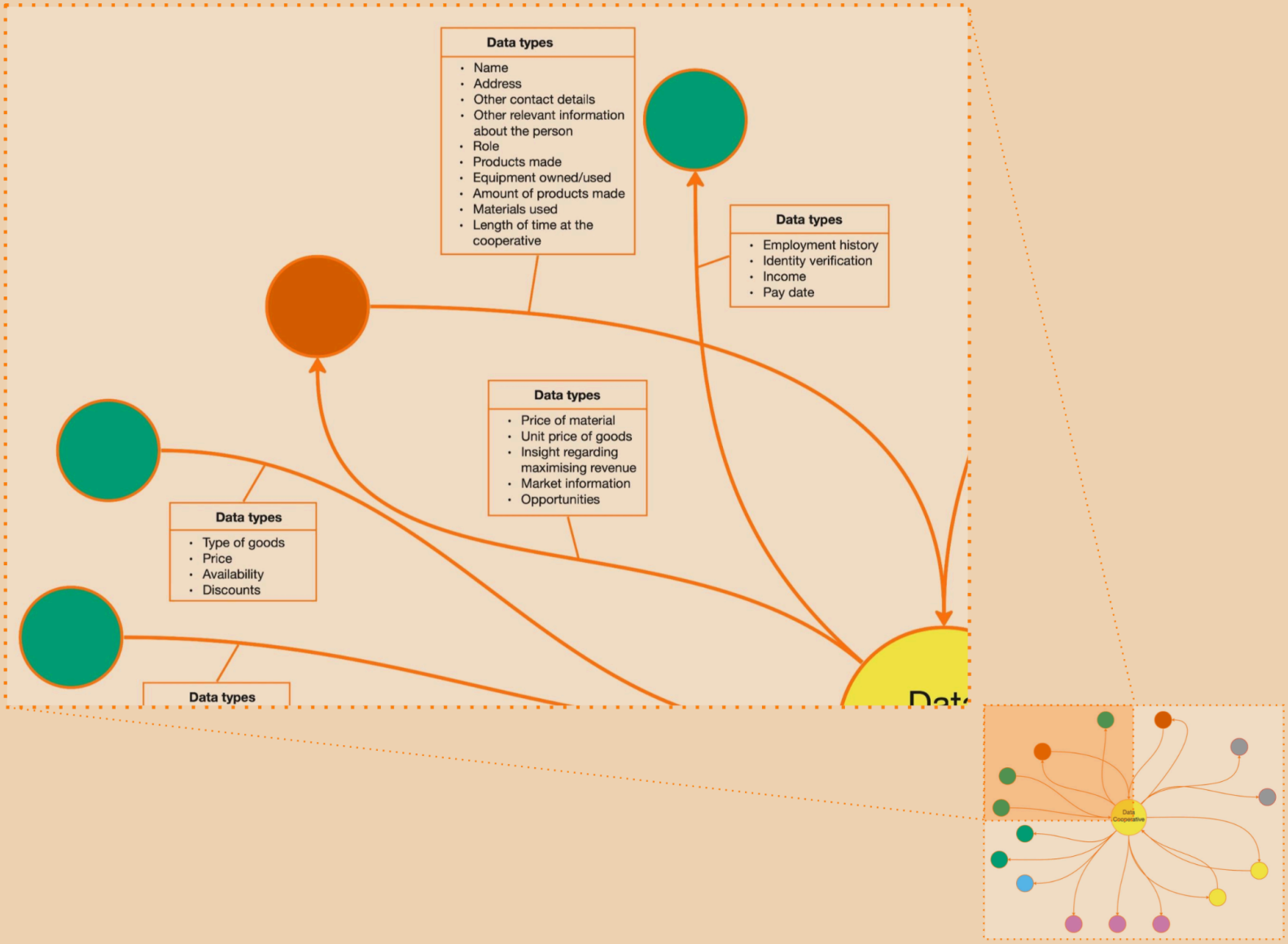
3. **Federated data flow** (Figure 4.3) is the sharing of information between organisations with the same aims, values, and data governance processes—especially cooperatives following the key principles outlined earlier in Box A. This type of data flow could enable small data cooperatives to work together.

4. **Third-party flow** (Figure 4.4) is a more traditional data-sharing or licensing agreement, where terms of sharing are agreed upon in formal contracts with third parties such as commercial organisations, government, or academic researchers. This could be done in exchange for services or money, or to meet regulatory compliance or research objectives.

One approach is to start by mapping primary data flows: direct transfer of data between the cooperative and key stakeholders. Secondary data flows, which show transfer of data from a key stakeholder to a third party, could then be added. Note also that the map should include both existing and potential data flows.

The next step is to integrate the data inventory with the stakeholder and data flow mapping. This can be done by assigning the types of data inventoried to every flow in the map (Figure 5).

Figure 5. Integrating data inventory and stakeholder/data flows mapping. Inset from Figure 3.



At this stage, the cooperative has a rough sense of the inflows and outflows of potential data exchanges, and is in a position to consider in more detail the value to all stakeholders—including the cooperative itself—of participating in such exchanges.

2.6 Hone in on the value proposition.

The cooperative is ready to identify the value proposition associated with managing and sharing pooled data. It's important to recognize that 'value' has different meanings. A conceptual discussion of the meaning of value (Box E) can help frame the more concrete conversation about the value a data cooperative provides to members.

Box E: What is value?

Value can be defined as that which is derived when an *interest* is fulfilled. Interests are anything that an individual desires—financial security or social equity, for example. These broad desires can be broken down into tangible benefits like access to credit or more environmentally sustainable product mix. Discussing interests helps members identify the value propositions of greatest importance.

Value has two components: content and context[23]. Content refers to the information contained in the dataset. Context pertains to the circumstances under which the data is gathered, analysed, and used. The economic and social value derived from data is dependent foremost on content, but context will influence the manner and extent to which the value is realised—and thus interests fulfilled. The value proposition conversation should consider the ability of the cooperative to control data content (actively working to improve the quality of data, for example) as well as influence context (working with other cooperatives to improve interoperability and thus usability of data, for example).

Another consideration when thinking about the value of data is that, unlike many physical goods, data can be shared or re-used without diminishing the available supply. Sharing imposes little marginal cost beyond the price of data transmission; most of the cost is associated with collecting, processing, and—for very large datasets—storing data. This means that cooperatives should think carefully about what rights they choose to retain or relinquish when sharing data.

Insofar as possible, a value should be ascribed to each data flow identified in the previous section's mapping exercise. This will help the cooperative and its members to prioritise the data they feel is most important and therefore should invest in managing. The valuation can initially be very approximate—for example, simple 'high, medium, low' value categories. It is important to remember that the cooperative will have legal and contractual obligations to keep certain types of data and, additionally, certain types of data might be needed for operational reasons. Such datasets could be kept separate from the valuation exercise.

Note also that the stakeholder and data flows mapping outputs should be treated as living documents. As time progresses, the cooperative may locate gaps in data flows or find new opportunities for relationships with data user stakeholders.

It's critical to keep in mind that value encompasses more than monetary returns. One possible way of assigning value to datasets is to borrow from business management theory's concept of the 'Triple Bottom Line' that considers economic, environmental, and social benefits. A simplified valuation matrix based upon the Triple Bottom Line can help evaluate the total value of each dataset (Table 2).

Table 2. A Triple Bottom Line valuation matrix.

Dataset	Economic value (1-10)	Social value (1-8)	Environmental value (1-6)	Total value (1-24)
Price of materials	8	5	1	**14**
Price of goods sold	10	5	1	**16**
Market information	8	3	1	**12**
Opportunities	5	7	2	**14**

There is a great deal of subjectivity involved in this exercise; comparing the relative value of different datasets across different notions of value is

challenging. The co-design workshop may seek to engage the cooperative's members in a more detailed conversation around valuation. A two-step process might entail:

1. Ask members to assign weight to the three 'value categories'. The facilitator can set a total number of 'value points' to all categories combined, and then ask members to participate in deciding how to divide the total points between the three categories. This division should reflect the overall purpose and principles of the cooperative, as well as individual member needs. In Figure 5's example, the total number of value points each dataset can obtain is 24, of which 10 are assigned to economic value, 8 to social value, and 6 to environmental value. In the example, the 'Unit price of goods sold' dataset attains the highest total value, 16 out of 24 possible value points.

2. Ask members to assign an economic, social, and environmental value to each dataset. The process of doing this may lead to a reconsideration of step 1, as well as a gradual and iterative adjustment of how to think about different point values within each category. In the above example, the 'Unit price of goods sold' dataset has the highest overall (16) and economic (10) value, but the 'opportunities' dataset has the highest social (5) and environmental (2) value.

This exercise may be time-consuming, but is worthwhile. In addition to more clearly identifying the value proposition, the cooperative will have a more nuanced understanding of the trade-offs and scope of data governance. The process suggested above should be refined to support the needs of each cooperative.

With respect to economic value, note that the granularity of a dataset can affect its value to data users. Broadly speaking, less anonymous, more disaggregated data is often more valuable, but comes with more privacy risks. These trade-offs will become apparent during the value proposition discussions, and can be discussed in greater detail at later stages of the

co-design process. Those subsequent conversations may lead to a revision of specific valuations and the value proposition generally. This kind of iterative work is both inevitable and valuable. Box F gives an example of one platform that generates economic value by pooling data.

Box F: JoinData

JoinData[24] is an agricultural data cooperative of Dutch banks, digital finance providers, and agribusinesses. Farmers pay a fixed amount of 50 euros per year for the use of the tool, which enables them to control their data through managing sharing authorizations. Farmers also play an advisory role in the cooperative's decision-making. Companies and other parties interested in using data generated by farmers pay JoinData a fee to enable the transport of data. Joindata recognises itself as a non-profit data cooperative wherein any revenues generated are reinvested directly into the platform, for example by improving the user interface.

Finally, note that value propositions may not fit neatly into traditional 'economic, social, environmental' categories, especially when considering new forms of digital data. Table 3 below lists examples of other ways tin which data cooperatives can facilitate the unlocking of data value, as well as possible long-term impacts.

Table 3. Data cooperative value propositions.

Value proposition	Long-term impact
Giving members the power to make decisions regarding how their data will be treated, whom it will be shared with, and how it will be used	Alignment of member decisions with member economic and social goals and principles
Access to credit and insurance by aggregating financial, etc. data and intermediating between the cooperative and financial service providers	Expansion of the activities of the cooperative and improved standard of living for members
Access to public sector benefits by storing and managing consent to personal data	Improved member well-being through participation in social welfare

Value proposition	Long-term impact
	schemes
Transparency from aggregating and sharing member data (with consent) to other co-op members	Trust between members
Accountability in governance of the cooperative using data on operations, finances, and management	Trust in the leadership of the cooperative
Better management of inventory and capital goods by aggregating data and providing insights on estimated requirements for a given output level	Operational efficiency
Better management of finances by aggregating data on revenues and expenditures	Increased risk appetite and greater willingness to adopt innovations
Maintaining best-practices standards of security and assuming responsibility in the event of breach	Greater member trust in data governance systems

Step 3. Design the governance architecture.

3.1 Assess and build trust.

The cooperative will succeed or fail based on the strength of its members' relationships. Well-constructed legal and technical frameworks can reinforce trust. Specifically, the horizontal decision-making structure of cooperatives, with legally prescribed roles for cooperative board members, is conducive to building trust. However, internal power dynamics and the motivations of individual leaders will also influence the outcome, and so even cooperative decision-making structures must stay vigilant about abuses of power.

In addition, even well-functioning formal mechanisms cannot be the sole basis of trust. Rather, deep and enduring trust is slowly negotiated over time in informal ways, especially as members observe the cooperative's willingness to prioritise transparency in all aspects of data management and sharing. This means making decisions with the participation and informed consent of its members, as well as meeting its stated commitments in returning value to members.

Overall, accountability (discussed at greater length in Section 3.2) is a watchword: members of the data cooperative should be able to hold the cooperative's leadership to account for all decisions, especially with regard to how personal data is handled. A foundation of trust also simplifies future decision-making. As members gain confidence in the cooperative's leadership to handle data responsibly, the desire to require case-by-case consent for all decisions may decrease.

Note that trust depends to a great degree on data literacy. For structures to be made fair, individual capacities must be supported and realised. Cooperative members must feel that they have understanding and agency over their data; trust is fundamentally about agency and having

mechanisms in place to support agency. However, even in the absence of sufficient levels of data literacy, the data cooperative must accept the responsibility of raising awareness among all members about all major decisions.

Because many data co-op members have experiences with marginalisation and over-surveillance, representation of vulnerable groups on governing boards is essential. The co-op must be wary of more data-literate members dominating decision-making processes. Community representatives may have less technical knowledge, but often have greater insight into operationalizing the values that drive the mission of data cooperatives—equality and justice, for example. Assuring that representation is more than tokenistic can be difficult. Cooperatives must strive to ensure that member voices are not only at the table, but are also driving the conversation, which often means careful planning around how conversations are facilitated.

Trust in the overall data ecosystem[25] in which the cooperative participates, not just internal trust within the cooperative, is also critical. Weak society-wide regulatory institutions lead to abuses of privacy and a resultant decrease in organisational trust. Lack of private sector transparency with respect to how data is used, for example in training decision-making algorithms, adds to the mistrust. Data cooperatives have the ability to counter these trends and build trust by showcasing transparency and providing individuals meaningful control over data. In fact, the data cooperative model may be uniquely valuable in helping communities disempowered through technology to regain their voice in the larger data economy, as well as become more visible within policymaking. Data can be, and has been, used as a form of control over marginalised people, but can also be an instrument to effect change.

There is no universal recipe for building trust, but the principles discussed throughout this handbook can guide data cooperatives: create strong, democratic formal mechanisms that stay vigilant of internal power dynamics and abuses of power; prioritise transparency and accountability;

include everyone in the conversation, including members who are less data-literate, and work to build data literacy; and participate in society-wide efforts to promote robust data regulation that gives power to individuals.

3.2 Outline an accountability and transparency strategy.

Given differences in members' data literacy and familiarity with data architecture, transparency is critical to data co-ops. At a minimum, transparency implies regular reporting about, and auditing of, how data is protected and shared. Optimally, transparency means achieving a universally inclusive, user-friendly, and efficient information flow.

Accountability and transparency should be 1) embedded in the cooperative's technological infrastructure, 2) reflected in the cooperative's organisational structure (e.g., committees), and 3) formally written into by-laws. Direct, automated reporting to members of how data is used is an example of a technology-mediated solution. An organisational structure approach would be to appoint an internal review board or officer(s) whose duty is to collect member feedback. The co-op's bylaws can function not only as a tool for compliance with broader legal requirements, but also as a means to describe internal processes and standards for member review. Overall, when accountability and transparency are included within the design and governance of a cooperative in diverse ways, trust will flourish.

Note also that balancing transparency with information overload is a challenge. Disclosures for transparency must be easy to understand and coupled with education programs for members at all digital literacy levels. Implementing such measures requires time and resources, but is essential for the long-term successful governance of the cooperative.

3.3 Consider the particular needs of data governance.

The cooperative will need to put into place systems and processes to handle data responsibly through its full life cycle, from creation to use to eventual destruction. This is the topic of data governance, which can be

defined[26] as ‘everything you do to ensure data is secure, private, accurate, available, and usable. It includes the actions people must take, the processes they must follow, and the technology that supports them’.

Responsible data governance ensures that the data collected is relevant to commonly agreed purposes, of high quality, adheres to data protection regulations, and reflects the data sharing preferences of members. Data governance processes also support regulatory compliance and risk management efforts associated with the data, particularly if the collected data is sensitive.

Within certain jurisdictions, when a cooperative manages personal information it takes on the role of ‘data controller[27]’, with legal responsibility for data collection, processing, storage, and sharing. Meeting these formal legal guidelines around data will require additional time and resources, beyond the more general costs of governance.

While a board or selected individuals may take on the role of setting up data governance systems, all members of the cooperative—regardless of data literacy—need to have a working understanding of these processes to effectively engage in decision-making around their own data. One strategy to promote this is to pair together members with disparate levels of data expertise to share knowledge and mutually check understanding of how the governance system functions.

Data is also a unique type of asset in the sense that it is held in disparate formats across multiple information technology systems and digital ‘filing cabinets.’ This distributed nature of data can make it more difficult for owners to have a comprehensive sense of their assets. The data cooperative must take on the responsibility of ensuring that owner visibility on assets is clear and available on-demand.

Note also that, as a general rule, data cooperatives should grant members extensive rights around their personal data, especially with respect to sensitive topics such as health, gender, and finances. Data cooperatives

must see relevant industry standards as the *minimum* threshold for data protection. Box G lists some commonly protected data rights.

Box G: Data Rights

General data laws offer cooperative members rights over their data. Not all jurisdictions grant each of these rights, but many bodies of law include:

- Right to be informed about data practices
- Right of access to personal data
- Right to correct or amend data
- Right to port data
- Right to appoint a personal data heir
- Right to restrict processing of data
- Right to withdraw consent over data
- Rights related to automated decision-making and processing
- Special rights protecting children

3.4 Decide on a consent mechanism.

The consent mechanism—the manner by which individuals decide how specific data is managed—is among the most important design decisions the cooperative will make. The frequency and granularity of consent that individuals prefer will differ based on who is using the data and for what purpose.

Consent plays a central role in the operation of data cooperatives from both a theoretical and legal perspective. Theoretically, seeking informed consent from members is consistent with the very purpose of data cooperatives: empowering data producers to have control over how their data is used. Legally, consent is increasingly recognized as a necessary prerequisite for the processing of personal data. Article 7, Recital 32[28] of the European Union's General Data Protection Regulation (GDPR), for example, suggests that authentic consent must be:

- *Freely given* without pressure or influence, and representing real choice;
- *Specific* about the purposes of processing and which data is involved;
- *Informed* as to the controller's identity, what data will be processed, how, and for what purpose, as well as informed about the right to withdraw consent at any time; and
- *Unambiguous* in the sense of requiring a statement or clear affirmative act.

While the above list tends to describe core features of how consent is defined in many jurisdictions, specific national standards may differ. The data cooperative should familiarise itself with the relevant legislation that defines a minimum standard for its consent mechanism.

A few additional principles are foundational to a consent mechanism. First, as noted earlier, data cooperatives should actively raise the skill levels and understanding of the data subjects—the member-owners of the cooperative—so that informed consent can be given freely. Second, regardless of the precise design of the consent mechanism and regardless of data literacy, all individuals should have the right to know to what end their data is being used, and all individuals should have access to the data they have shared. Third, consent should be seen as part of a trust-building process. Consent is about the ability to empower members to share and revoke their data freely. When cooperatives use legally valid consent mechanisms, trust becomes embedded in the structure of the cooperative.

While no one-size-fits-all consent mechanism design exists, there are three generally recognised models[29]. When choosing among these, data cooperatives should consider how each model affects efficiency, individual control, and data sharing, as well as aligns with the cooperative's values.

1. **Granular consent**: Individuals explicitly and frequently consent to data collection and sharing, similar to cookie preferences on websites. Every type of data is subject to this decision, and members are additionally able to customise with whom they share data and in what form. Consent can be withdrawn at any time.

2. **Persona/archetype consent**: Consent is based upon the characteristics of a chosen persona or archetype. Each persona is associated with a set of data sharing values and consent guidelines. Individual members select personas whose values and broad consent preferences match their own. These personas automate sharing decisions across datasets; individuals do not have to participate in every data sharing decision.

3. **Traffic light consent**: This model offers three data sharing options: green, amber, and red. The ‘green’ option is most permissive, allowing the cooperative to share the members’ data how it sees fit. The ‘amber’ setting allows individuals to opt for granular consent for specific datasets while retaining control of more sensitive data such as names or addresses. The restricted ‘red’ option allows the data cooperative to use the chosen data only for its own internal purposes. The specific permissions granted by each colour can be customised according to the needs of the cooperative.

Each of the three models comes with advantages and disadvantages. Granular consent offers the most control and transparency to individuals; choices are highly customisable. However, granular consent relies on people having the time to understand and evaluate the need for consent frequently, possibly in each data sharing case. If time is not available, people may either make poor choices, accept the ‘default’ option, or not make a choice at all. For the cooperative as a whole, granular consent can lead to inefficiency. If each data-sharing request is considered by each individual, the smooth flow of data to end users is greatly slowed down.

The persona/archetype model, on the other hand, can be highly efficient. Because any future data would be shared on the basis of agreed and

understood value sets, persona models reduce the burden of consent choices on individuals and allows cooperatives to implement rapid data flows. The trade-off is loss of control. Once a persona or archetype has been chosen, the onus is on the individual to check regularly to ensure that the initial choice corresponds to their evolving knowledge and values. In addition, the cooperative must create and revise personas transparently to ensure that the offered personas do not overgeneralize the member base or reinforce harmful biases about member value sets.

The traffic light model is the most flexible, as individuals can select different colours for different data or end user types. While the traffic light model places fewer demands on data literacy and time than the granular consent model—particularly when 'green' and 'red' permissions are chosen—it does require individuals to discern how their preferences vary across data types and end users. The effort spent making choices is likely to be greater than in the persona/archetype model.

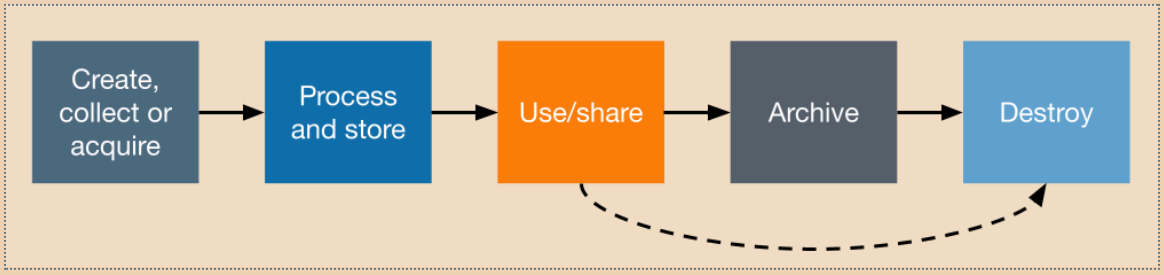
Although describing and discussing the consent mechanism may require significant time during the co-design process, the investment is valuable. Fundamentally, the consent mechanism conversation is about the distribution and delegation of power within the data cooperative. Cooperatives are based on the principle of 'flat' decision-making, wherein every member has equal voice. However, nearly all cooperatives delegate duties to officers, some of whom are directly elected—as in a Board of Directors—and others who are hired by the elected leadership. Choosing a consent mechanism is similarly about striking a balance between democratic expression and delegation. Individuals can value the same data differently, and although all members might be aligned to the cooperative's mission, the reasons for that alignment may differ—and the differences will be reflected in choices about the consent mechanism.

Even in cases where members have formally delegated their decision-making authority to others, the cooperative's leadership should keep in mind the trade-offs between the value of sharing data and helping members regain and maintain control over their data. An internal ethical

review board may help in reviewing data requests; for smaller co-ops, third-party support from academic institutions and civil society, as well as established data cooperatives and their support organisations, may be more feasible. Member engagement and communications campaigns can serve to present the opportunities and risks of consent, including clarifying legal rights around withdrawal of consent. Overall, the cooperative must create an environment of trust in which individual members know that their rights extend beyond the lawful consent process, while also acting as a viable partner to potential end users of data. A more trusting environment will also likely lead to more stable consent decisions by individuals.

Note also that data governance covers all aspects of the data lifecycle (Figure 6), from collection to destruction. Consent mechanisms should encompass all steps of this lifecycle, and additional architecture may be necessary to support each step. For example, user authentication is a first step to ensuring that the data subject themselves is the one consenting.[30] Especially (but not only) in granular consent systems, significant technical infrastructure and capacity is necessary to manage consent.

Figure 6. The data lifecycle. Note that data may be destroyed without being archived.



Finally, cooperatives should be wary of ‘consent fatigue’ among members, which undermines their ability to make informed decisions about their data. This is where consent mechanisms must strive to be inclusive of diverse capabilities, data literacy, and available time. The cooperative does not simply exist to ‘arrive at consent.’ Rather, it’s a vehicle for guaranteeing the rights and promoting the autonomy and well-being of its

members, which means that the process of deciding consent is as important as the final decision itself.

3.5 Plan for scaling up.

Data cooperatives may require a minimum size to be viable. There are costs to establishing and maintaining a data architecture, and datasets need to be sufficiently large to have economic value. A plan for growth and scale is therefore important, but should be balanced with maintaining the co-op's values. For some co-ops, staying small and better aligned with the needs of a well-defined community is the best path, regardless of the benefits of scale. Maintaining a smaller scale is possible, but small data cooperatives may have to rely on subscription or worker-equity models in which members contribute their money or labour to maintain the co-op, and/or seek grant funding.

Cooperatives that do plan to scale should proactively consider how the value of their data will grow as datasets become larger and attract the interest of external organisations. Co-ops may consider different pathways of data access for different external organisations, based on benefit to the co-ops members, alignment of values between organisations, and other considerations. At the formation of the data cooperative, members may want to consider inserting an 'asset-lock[31]' or an anti-demutualisation[32] clause into their by-laws to act as a legal block against predatory behaviour by external organisations.

Box H describes the Mondragon Cooperative, a successful example of cooperative scaling.

Box H: The Mondragon Cooperative

The Mondragon Cooperative[33] is a remarkable example of successful scaling of cooperative structures. It is currently an association of 95 autonomous cooperatives in which each worker is a member-owner and votes on vital decisions within each cooperative. Profits are shared amongst members of these individual cooperatives.

In case of losses, both funds and workers are shared across the network of cooperatives. All cooperatives share detailed information on earnings and other key topics. Workers buy into the co-ops with one-time payments, which are stored as funds that can earn interest and with a portion of dividends added to the balance. Mondragon's network spans many different sectors and in 2021 brought in more than eleven billion euros in revenue. Mondragon's experience suggests the potential of cooperatives as a viable, profitable, and sustainable alternative to conventional businesses.

Scaling must not be done without careful planning, however. At a minimum, the following questions must be discussed by the cooperative's members:

1. Is scaling necessary for the data cooperative to achieve its objectives?
2. Is size a critical indicator of a data cooperative's health or efficiency, as the members themselves define these and other goals?
3. What risks does scaling pose, and how can these risks be mitigated?
4. Does scaling up contribute to the co-op's long-term sustainability?

The last question about sustainability is critical. Cooperative principles intrinsically embody sustainability[34], especially the social aspects of sustainability. Some key determinants of sustainability include:

- *A sense of member ownership:* Cooperatives are usually formed to address issues that are hyper-local in nature. Thus creating a sense of ownership in the data cooperative within the socio-cultural context is extremely important to build trust and confidence, and thus to improve long-term sustainability of the co-op. Inclusive co-design processes help members identify as partners rather than mere passive data collectors.

- *Equitable revenue allocation schemes:* Designing fair and equitable revenue allocation mechanisms is critical to ensuring the sustainability of data cooperatives, and influences the quality and quantity of data that they share; Swash[35] is an example of a cooperative that has thought carefully about this issue. In many cases, the data cooperative should focus on creating incentives for members to participate and share high-quality data which, when aggregated, can have an extremely high market value, thus increasing the cooperative's revenue. When incentives are well-aligned with individual goals, members tend to be more committed to the work of the cooperative. However, some research indicates that distributing revenue proportional to the data contribution[36] may not be optimal for the long-term sustainability of a data cooperative. In some cases, it may be better for the cooperative's leadership to decide profit allocation based on criteria other than data contributions.

- *Identifying independent streams of revenue:* Initially, a data cooperative may be dependent on grants or angel investors. However, data cooperatives must quickly find independent[37] streams of revenue so that they are not co-opted by technology companies or state bodies, and continue to serve and protect the rights of members. Put simply, the long-term sustainability of a data cooperative depends critically on the relationship of earned revenue to operating costs.

- *Overcoming challenges posed by heterogeneity:* Aligning the goals of all members is a complex task, especially in larger cooperatives[38]. With size comes accountability for larger amounts of data, making it harder for leadership to make decisions that satisfy all individuals. For example, Driver's Seat Cooperative was founded in 2019 to help collectivise the gig economy.[39] However, in the ride-hailing business model, drivers not only have shared interests but are also competing for rides in the same area, and so needed assurance that

the algorithms which affect their employment were not using, even unintentionally, discriminatory practices that favoured others.

- *Maintaining trust:* It is important to remember that scale may lead to an erosion of shared values, as well as more centralised decision-making to maintain nimbleness in the face of greater organisational complexity. This can make preserving trust difficult, especially as new members and other stakeholders enter. The challenges and costs of being transparent also increase with scale, so transparency policies will also need to be adapted. Additional accountability measures may also be necessary to ensure that members have a say about changing levels of transparency. Because of these considerations, it may be wise to layer data co-ops on top of existing offline co-ops that already have robust systems of governance at scale.

Step 4. Create a data infrastructure.

By way of introduction to this section, we note that striking the right balance between collective and individual needs is critical to an effective data infrastructure. Data is by nature relational and networked: the most useful insights emerge when individual data is pooled and shared[40], revealing relationships and interactions. However, data is currently often conceptualised solely as private property, and as a result many datasets currently sit in silos, which inhibits communities and organisations from maximising potential value. Designing data infrastructures for security and safe sharing must be thus done in a community context, not only by individuals making personal decisions.

However, even within an organisation with well-defined common values, individual needs and preferences do vary. Designing individual incentive mechanisms to promote data sharing is thus important. Identifying the priorities and goals of individuals during the co-design process can highlight the different types of incentive mechanisms that may be effective in encouraging data sharing. One prominent example is MiDATA[41], a health data cooperative founded in Zurich, Switzerland, which creates common value by allowing individual users to share data for specific medical research projects.

More broadly, the FAIR (findability, accessibility, interoperability, and reuse) principles[42] for data management and stewardship are useful guidance to keep in mind when creating a data infrastructure.

4.1 Conduct a risk assessment.

The first step is to conduct a risk assessment. Organisations often over-collect data that, when assessed, has little or no value. Apart from creating an administrative and technical cost to the cooperative, if the data contained is of a sensitive nature, then it might carry extra risks that outweigh potential gain. A risk assessment process allows the cooperative

to make decisions on unnecessary datasets—for example, minimising the types of information held within the dataset, stopping collection of data for the dataset, or even deleting the existing dataset altogether.

Conversely, organisations may miss opportunities to utilise data. The risk assessment process offers a chance for the cooperative to identify instances in which data could be collected and shared without exposing its members to undue risk.

Some of the data that the cooperative collects will be sensitive because it contains personal, commercial, or other types of private data. In most jurisdictions, personal data is protected by statute. For instance, within the United Kingdom's adopted version of the European Union's General Data Protection Regulation, called UK GDPR, the following categories of data are given greater statutory protections and need extra care when processing: personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership; genetic data; biometric data used for identification; data concerning health, a person's sex life, or a person's sexual orientation.

Such datasets should be singled out as priorities for risk assessment and specific strategies devised to reduce the chances of re-identification or reconstruction of the data. If the sharing of sensitive data is a key revenue strategy for the cooperative, such as in a health data cooperative, then appropriate safeguards should be implemented—security measures to minimise the likelihood of unauthorised access and use, as well as verified logging of data sharing and use. In addition, processing measures such as aggregation, anonymisation, pseudonymisation, or synthetic data generation can help reduce risk, as we discuss further in Section 4.4.

Box I describes the creation of a 'data risk register' to carry out a risk assessment.

Box I: Data risk registers

Creating a data risk register will help those responsible for the security and confidentiality of data to have oversight and control of how data is managed. One easy way to develop and manage a risk register is to create a spreadsheet with rows listing each specific dataset risk and columns capturing the variables described below.

- *Risk number:* unique identifier of the risk.
- *Data asset*: unique identifier of the dataset to which the risk pertains.
- *Owner*: the person(s) responsible for managing the risk.
- *Description of risk*: a full description of the concern (e.g., access to data from unauthorised people, people copying data to other devices, etc.).
- *Potential impact (qualitative)*: the consequences of the risk being exploited. Impacts can be financial, operational, reputational, member-focused, etc., given the specific context of the cooperative).
- *Potential impact (quantitative)*: translation of the impact into a numerical scale. An example of a simple scale that could be used: {1. Trivial; 2. Minor; 3. Moderate; 4. Major; 5. Catastrophic}.
- *Probability of impact*: estimation of how likely a risk is to occur. An example of a simple scale: {1. Almost impossible; 2. Unlikely; 3. Possible; 4. Likely, 5. Almost certain}.
- *Overall risk rating*: multiply the potential impact and probability impact scores to obtain an overall risk rating for the specific risk in question. The overall risk rating could then be interpreted through creating 'action priority' categories, e.g., {1-3. Very low risk, action optional; 4-6. Low risk, should be addressed where/when possible; 8-12. Medium risk, requires action; 15-16. High risk, requires priority action, should be escalated; 20-25. Very high risk, requires immediate action}.

Risk Rating						
Impact	5	5	10	15	20	25
	4	5	8	12	16	20
	3	3	6	9	12	15
	2	2	4	6	8	10
	1	1	2	3	4	5
		1	2	3	4	5
		Probability				

Legend	
	Very low risk, action optional
	Low risk, address where/when possible
	Medium risk, requires action
	High risk, requires priority action, escalate
	Very high risk, requires immediate action

- *Current risk mitigation actions*: the current actions and strategies being taken to mitigate the risk.
- *Proposed risk control actions:* future plans to improve the situation and reduce the risk rating.

4.2 Devise a data collection strategy.

The tools that the cooperative uses to collect data should be appropriate to its available resources, including the capacities and needs of its members and the funds available for digital data management. Regardless of the specific informational technology solutions chosen, data should ultimately reside in a database subject to the data governance processes of the cooperative.

We can categorise data collection processes as either manual, in which someone observing or acting is part of the data collection process, or automatic. Each option can be carried out through various means, but it is important that the data collection workflow is standardised in all cases, and that it is understood by those involved, relevant to the task at hand, and not unnecessarily burdensome for data-holders or data collectors.

If the process is manual, enumerators should be trained to ensure that they follow standardised processes for data collection. They should also be provided appropriate resources, including prepared templates or

forms, so that data collected by different enumerators is comparable. Manual data collection can happen either through digital tools (see Box J for an example) or paper-based methods, which can be effective in contexts with limited digital infrastructure and technical resources. If using paper-based methods, it is important to ensure that there is a standardised way of capturing and coding this information into digital form later.

Box J: Our Streets Chorlton data gatherer

In the Our Streets Chorlton[43] project, in Manchester, UK, groups of local residents who were concerned about the impact of traffic on children attending local schools used a simple tablet and smartphone tally counter to register the volume and types of traffic that passed local schools during the day. The tally counter was configurable, logged the time when the vehicle was counted, and allowed the data to be exported as a comma separated value (CSV) file that could be easily be imported into spreadsheet programs. This data was collated, analysed, and presented to policymakers.

Automated data collection is also becoming easier. Low-cost technologies and almost ubiquitous mobile phone connectivity have created an explosion in the availability of low-cost sensors that can detect and measure different phenomena. For example, ride-share, delivery, and satellite navigation apps all harness smartphone capabilities, especially global positioning system (GPS) features and wireless internet access, to record and transmit location and mobility data.

The data from sensors is often sent back to a server which allows the data to be written to a database automatically. One major factor to consider when using mobile phone data and other automated data sources, however, is the issue of consent. Mobile phones have a wealth of automatically collected data which can be extremely useful when used appropriately. Though the data is automatically available and requires little

human effort to collect, consent to sharing and using the data from users is a crucial step in the data collection process.

Overall, digital tools can allow data to be collected in a more structured and coherent way, and can create efficiencies as the data generally does not have to be re-entered later. However, sometimes it is more appropriate to use more manual methods of collection. Typically, gathering data manually requires less technical expertise, and is often better suited to secure granular consent, for example when taking photographs or making video and sound recordings.

4.3 Create a secure, trustworthy storage system.

Solutions for storing the cooperative's data assets can be broadly split into two categories:

- On-premises storage: information technology facilities on your premises, possibly on the cooperative's own servers.
- Cloud-based storage: data storage on cloud storage providers off-premises.

Each category has its advantages and disadvantages, and there are many potential providers of servers and associated equipment for each option. To make the right decision for the cooperative, it is important to have completed the previous steps outlined in this handbook, especially with respect to understanding technical needs (what data is being collected, how, by who, for what purpose, and data flows across stakeholders) and answering key data governance questions (sensitivity of the data, how data is managed, access requirements).

With on-premises solutions, organisations must keep their systems up-to-date and maintain security. If lost or stolen equipment holds personal or sensitive data, then there may also be legal and contractual consequences, as well as potential harm to individuals and the cooperative if stolen data is used maliciously.

Cloud-based solutions allow organisations to host and share data from remotely located data centres. Data centres will usually have heightened levels of security with respect to both physical and virtual access, redundancy plans for technology in case of equipment failure, and ultra-high speed data connections to allow rapid transmission of data. In addition to data storage space, many cloud-based solutions providers offer Software as a Service (SaaS) tools which can be accessed through a secure web browser or local applications. The best cloud-based providers utilize current best practices in security, for example two-factor authentication.

However, good internet connectivity is needed for cloud-based solutions. Accessing files will be difficult without reliable connectivity, and many of the applications exist online. Note that connectivity can be an issue with on-premises servers as well, especially if they need to be accessed by third-party or remote clients. Connectivity also plays a role in the choice of data collection tools. Ideally, data collection applications or remote sensors would send data directly to the cooperative's servers or cloud-based data storage. If connectivity issues exist, however, data input might instead be done by manual entry into spreadsheets.

A major factor in deciding whether to use on-premises or cloud-based solutions is national data storage laws. Many countries have strict laws limiting the data storage options available to cooperatives. For example, when handling personal information, it is often necessary that the data is held within the country where it is collected, unless there are appropriate data adequacy arrangements that allow the transfer of data more widely.

Generally, if good internet connectivity is available, we recommend using cloud-based services, which have a number of advantages. Little data is held on local computers, so the likelihood of data breach or loss is minimised. Data held in a cloud database minimises versioning issues, in which people have different versions of the same dataset on their computers. Data and SaaS held in the cloud can generally be accessed

from anywhere connected to the internet, enabling remote work. Cloud-based solutions are also scalable, depending on evolving needs.

Cost considerations will also play a role in the cooperative's data infrastructure decisions. Generally, costs can be broken down as capital expenditure (CapEx) and operational expenditure (OpEx). Capital expenditure covers the IT equipment that the cooperative will purchase and own, such as computers, servers if hosted on premises, and any associated equipment. These items will have an upfront cost and, once purchased, lose their value over time. Operational costs cover the ongoing costs of running the cooperative. These can include wages and labour costs, equipment leases, cloud storage costs, software licences, electricity, rent, and internet connectivity. Data collection will also have costs associated with the equipment, software, and people needed to install, operate, and maintain the software. Some of these costs could be reduced by spreading some responsibilities on a volunteer or at-cost basis across the cooperative's membership.

4.4 Create safe, flexible ways of sharing data.

Being able to share and transact data is core to the cooperative's function. Sharing can take place within the organisation and/or with third parties.

Once consent has been given (see Section 3.4), the cooperative has two broad categories of options for allowing third-party access: using APIs, for example a dynamic dashboard that summarises key insights and analytics from the underlying databases, or allowing users to request and download data files themselves. Both routes involve some cost, whether it be the cost of building and maintaining the APIs so that they share data appropriately and securely, making the data files available for download on a webpage, or data requests being dealt with manually.

Being able to write, retrieve, and share data held on the cooperative's servers is reliant on connectivity and technical capacity, which also has its associated costs. Data collected via mobile phones may also incur data

charges from the network provider. Similarly, autonomous sensors can incur data transmission charges, depending on the technology used.

Legal requirements play a role in shaping the creation of a data sharing framework. Legal standards for data transfers typically fixate on the adequacy of the mode of transfer and the recipient's data protection standards. What constitutes 'adequacy' of safety and security are not usually defined explicitly, but are interpreted to conform to industry standards, as well as standards on par with a jurisdiction's own data security and privacy laws.

GDPR Article 32 suggests that adequate or 'appropriate' security might involve: pseudonymisation, encryption, and other privacy protection techniques (see Box K); the ability to ensure the ongoing integrity of processing systems; the ability to restore access in the event of a technical incident; and/or a process for regular security testing. Some jurisdictions such as India also have localization requirements for certain types of personal data, thus restricting cross-border transactions. Specific technical requirements for data sharing transactions should be detailed in contracts.

Box K: Privacy protection techniques

Techniques such as anonymisation, obfuscation, aggregation, and pseudonymisation can remove or severely restrict the ability of the data subject to be reidentified. Anonymisation[44] is the process by which all personal identifiers are removed from the dataset. Pseudonymisation[45] is the processing of personal data such that it cannot be traced back to the individual to which it belongs. Obfuscation[46] is a data masking method by which personal data values are replaced with values which match the original variable format. Aggregation[47] is the process by which raw data are gathered, reformatted, and presented in a summary form for subsequent data sharing and further analyses; the summary information cannot be traced back to individuals. The creation of purely synthetic data based upon sensitive data is another useful technique.

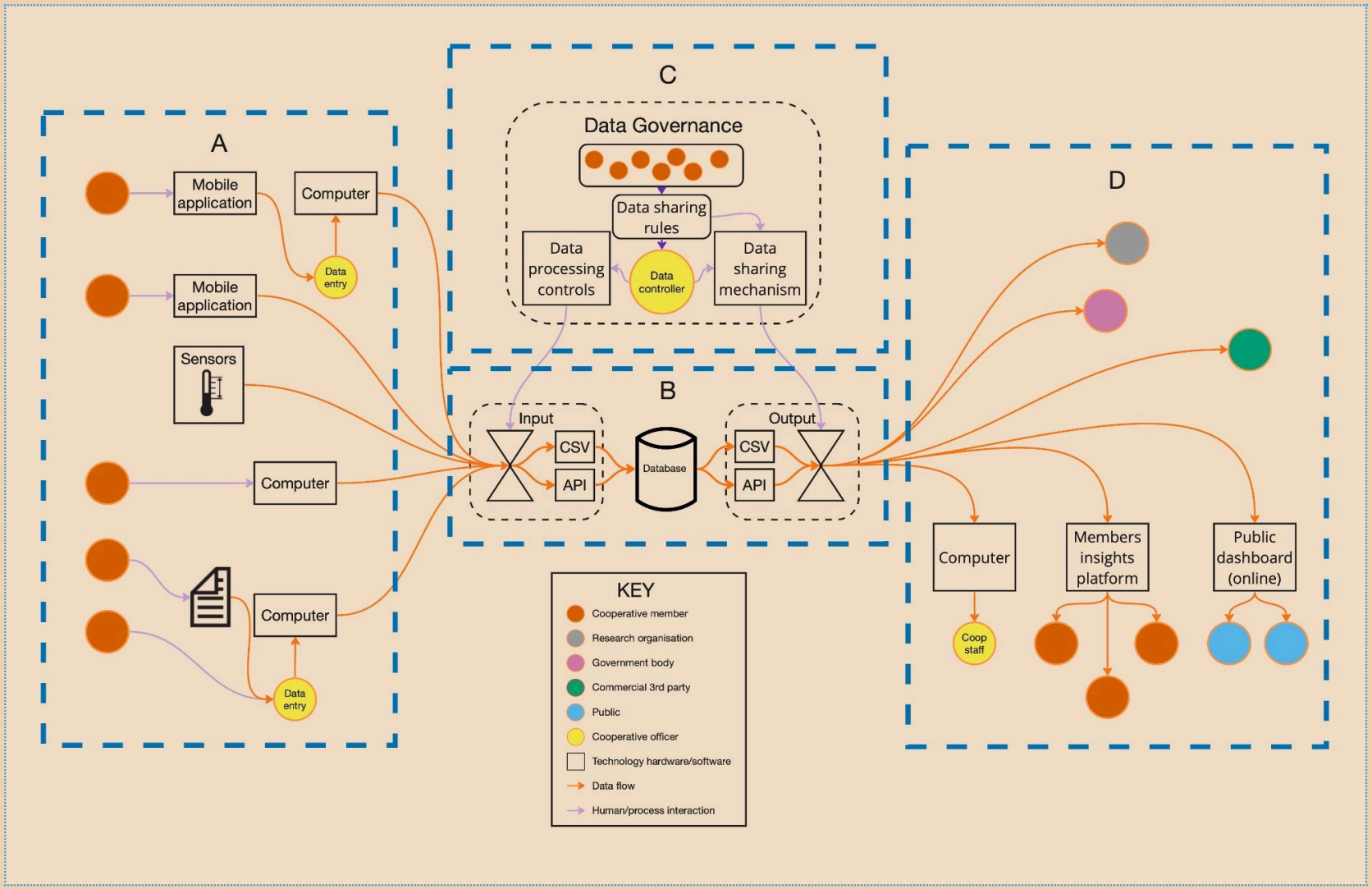
The cooperative's values should inform who within the organisation is responsible for managing data sharing transactions and which parties are eligible to receive datasets. When data sharing decisions are aligned with cooperative values, data subjects can better predict the ultimate uses for their data, making the consent process more efficient and extending trust from the member to the third party. For example, SalusCoop[48], a health data cooperative, has a consistent and specific data sharing framework based on criteria used in established hospital ethics committees. The cooperative limits transactions to non-lucrative and open science. By doing so, it allows for self-selecting on the part of third-party organisations who may already be familiar with such criteria, and it provides credibility and consistency for members.

With such models in mind, data cooperatives could appoint internal committees to review data sharing proposals and ensure their purposes are consistent with the cooperative's mission. Some jurisdictions even require the appointment of legally-required Data Protection Officers (or similarly named officers) to fulfil such a role. Even without such formal mechanisms, cooperatives can give members a degree of participation in the decision-making process, for example through focus groups that represent the larger member base.

Figure 7 presents a schematic of a generalised data cooperative data management and governance structure, summarizing the discussion in various sections of this handbook. Zone A shows the myriad ways in which data can be collected, from manual collection and inputting into the cooperative's database to automation via sensors and directly connected computer terminals. Zone B illustrates the technical data infrastructure that allows data collection, storing, and sharing. Data could be ingested to the cooperative's database via spreadsheets in CSV (or other) format or via an API (Application Programmable Interface), wherein a computer programme, mobile application, or remote sensor can write data directly to the database. Similarly, data could be shared directly to data users in the same ways. Zone C shows the governance process that not only decides which data is collected and stored and the mechanisms for doing

so, but also who has access to the data and under what conditions. These processes are determined by the data cooperative members themselves, or by nominated ‘data controllers’ to oversee this process. Zone D illustrates various ways of making data available to end users.

Figure 7. Generalised overview of data cooperative infrastructure. Zone A: ways in which data can be collected. Zone B: the technical infrastructure for data collection, storing, and sharing. Zone C: the governance process to make data management decisions. Zone D: Ways of making data available to end users.



4.5 Design for participation and inclusion.

At present, community participation in the collection, storage, use, processing, sharing, and disposal of data is rare. Further, the use of artificial intelligence (AI) and machine learning (ML) models has led to increasingly extractive[49] data practices, excluding communities from participating in their own data governance.

The emphasis on participation must extend beyond the co-design workshop. Involving communities in data collection is especially important. Training cooperative members to collect high quality data can help build datasets that would, for an outside organisation, be difficult to collect and verify. Norms and standards for data collection could be established through regular dialogue that deepens the culture of data cooperation.

Building inclusion will require time. Co-design of novel technological structures can be an arduous process; many individuals may reject new ideas and prefer the status quo. The cooperative should respect these decisions, but also work to address the constraints that feed into such attitudes—for example, poor internet connectivity, personal time constraints (including for learning technologies or attending workshops), and lack of trust in sharing data. 'Simplification' is a watchword for data cooperatives, not only in the co-design process but also with respect to the governance and sharing of data. Sophisticated tools are only appropriate if support resources exist and are accessible, and if the use of these tools is of clear value to the cooperative's members.

Finally, inclusion of stakeholders outside the cooperative is also essential. For example, collaborative roadmapping sessions with software companies can ensure that the code development is aligned with stakeholder needs, preventing the need to spend time and resources to later modify the software. Generally, promoting a culture of open modification of code promotes a sense of co-ownership throughout the

process; code should, as much as possible, be open-source and accompanied by comprehensive and user-friendly documentation.

Conclusion: Legal considerations

The previous chapters have outlined the key steps in building a data cooperative: confirming that the model is right for your needs; identifying the value proposition; designing a governance architecture; and creating a safe data infrastructure. Once the co-design process is complete, your organisation should analyse the legal context around establishing and operating a data cooperative. This concluding section discusses some broad considerations. More detailed resources are available[50] for further guidance.

Globally, data cooperative law is in a nascent stage. The landscape is evolving—various jurisdictions are experimenting with legal guidance around new forms of data governance—but to our knowledge, at the time of this writing, there is no country which has laws that exclusively govern data cooperatives.

Data co-ops must thus look to laws which govern other areas, but may nevertheless affect data cooperatives[51].Two bodies of law are especially important: cooperative law and data protection law. The first step before initiating the legal process of establishing a data cooperative is to ascertain whether these bodies of law allow data cooperatives. Obtaining a legal opinion is advisable in this step. Legal guidelines relating to the following topics are especially relevant:

- Purposes for which cooperatives can be established
- Activities permitted or prohibited
- Mandatory principles governing cooperatives
- Allowable relationships between the cooperative and its members
- Provisions on structure, membership, and governance of cooperatives
- Requirements in relation to registration of cooperatives
- Subnational (e.g., state/province) laws affecting cooperatives
- Sector-specific laws (e.g., agriculture, health) affecting cooperatives
- Compatibility of data protection law and cooperatives law

Note that, in many countries, establishing a data cooperative will require certain legal formalities such as registration with a public authority, publication of by-laws, or disclosures to a data protection agency.

On top of any rights, duties, benefits, and obligations dictated by law, the data cooperative can establish by-laws to formally recognise additional rights and prescribe specific duties towards its members, the cooperative itself, and any third party dealing with the cooperative. The governance framework outlined in Step 3 should be formally articulated in the by-laws.

Note that existing organisations seeking to add a data cooperative layer may need to engage with specific additional legal provisions. One pathway is an 'exit to community' strategy[52] by which a company with an established legal structure—and a strong foundation of user trust—could be bought out by its users to take over and form a cooperative. However, this transition may come with complications; existing organisational practices, for example, can hinder the widespread adoption of new norms around data cooperation. In some cases, establishing new cooperatives from scratch may be a better approach, especially if the founding membership is strongly aligned on data governance values.[53]

Finally, since data cooperatives look after data on behalf of their members, their legal and ethical responsibilities in relation to data are equivalent to those of a fiduciary. Per the principles of corporate law, a fiduciary has the following duties towards its beneficiaries:

1. *Duty of care*, which requires that the organisation be administered competently and diligently so that the interests of the members are not harmed; and
2. *Duty of loyalty*, which provides that any action taken by the organisation be in the interest of the members.

Note that these duties need not be enumerated as strict legal or contractual responsibilities within official documents; they are assumed for any organisation playing the role of a fiduciary.

'Information fiduciaries[54]', by the nature of their relationship with the beneficiary, owe duties in relation to information. Lawyers, doctors, and accountants are traditionally grouped under this heading. Since data cooperatives have access to member data, they are also best categorised as information fiduciaries.

In short, data cooperatives must use the data they obtain from members only to the advantage of the members. The duties of care and loyalty should be embedded in the structure of the data cooperative through design and governance choices that foster empowerment, transparency, and accountability. Clear policies around the use of and access to member data help reaffirm fiduciary responsibilities.

We close this handbook by emphasising the importance of partnering with like-minded organisations. Leaning on the broader cooperative ecosystem can help reduce administrative and operational costs for data co-ops. This approach is also aligned with the principle of 'cooperation among cooperatives' by 'working together through local, national, regional, and international structures' to 'improve services, bolster local economies, and deal more effectively with social and community needs' (refer to Box A for a full list of the cooperative principles). One prominent example is a federation of farmer cooperatives in Uganda that has piloted a public-private consortium[55] that pools and harmonises data to improve decision-making among all stakeholders.

The success of data cooperatives ultimately hinges on the ability to join forces with others in the cooperative movement. This certainly includes other data cooperatives, but also traditional workers' cooperatives, consumer cooperatives, credit unions, and other allies. The future of data cooperatives is bright, and rests of the successful creation of ever-wider circles of solidarity.

Works cited

1. International Cooperative Alliance. Facts and figures. https://ica.coop/en/cooperatives/facts-and-figures (2024).

2. Dave Grace and Associates. *Measuring the Size and Scope of the Cooperative Economy: Results of the 2014 Global Census on Co-Operatives*. https://www.un.org/esa/socdev/documents/2014/coopsegm/grace.pdf (2014).

3. International Cooperative Alliance. Cooperative identity, values & principles. https://ica.coop/en/cooperatives/cooperative-identity (2024).

4. Data2X. *Big Data, Big Impact? Towards Gender-Sensitive Data Systems*. https://data2x.org/wp-content/uploads/2019/11/BigDataBigImpact-Report-WR.pdf (2019).

5. Higgins, S. Gender-differentiated credit scoring: a potential game-changer for women. *Financial Alliance for Women* https://financialallianceforwomen.org/news-events/gender-differentiated-credit-scoring-a-potential-game-changer-for-women/ (2020).

6. Open Data Manchester. *Developing a Data Cooperative Model for Small Energy Cooperatives*. https://raw.githubusercontent.com/OpenDataManchester/data-cooperatives/main/Data%20Cooperatives%20Report%20FV3G.pdf (2021).

7. Aapti Institute. Exploring the value of adding a data layer to cooperatives : Megha farmer cooperative case study. *Medium* https://aapti.medium.com/exploring-the-value-of-adding-a-data-layer-to-cooperatives-megha-farmer-cooperative-case-study-1c4fcfd08635 (2022).

8. Ada Lovelace Institute. *Exploring Legal Mechanisms for Data Stewardship*. https://www.adalovelaceinstitute.org/wp-content/uploads/2021/03/Legal-mechanisms-for-data-stewardship_report_Ada_AI-Council-2.pdf (2021).

9. Micheli, M. *et al. Mapping the Landscape of Data Intermediaries: Emerging Models for More Inclusive Data Governance*. (Publications Office of the European Union, Luxembourg, 2023).

10. Taylor, L. What is data justice? The case for connecting digital rights and freedoms globally. *Big Data Soc.* **4**, 2053951717736335 (2017).

11. Vyas, P. & Young, R. Redefining socially responsible designing to assist collaborative approaches to community engagement. in (Cambridge, UK, 2011).

12. Inclusive Design Research Centre. Community-led co-design kit. *Community-Led Co-design Kit* https://main--co-design.netlify.app/ (2024).

13. Kenny, U. & Regan, Á. Co-designing a smartphone app for and with farmers: Empathising with end-users' values and needs. *J. Rural Stud.* **82**, 148–160 (2021).

14. Wolff, A., Barker, M., Hudson, L. & Seffah, A. Supporting smart citizens: Design templates for co-designing data-intensive technologies. *Cities* **101**, 102695 (2020).

15. Vogel, S. Co-creating a more equitable world: the transformative benefits of participatory design. *MIT D-Lab* https://d-lab.mit.edu/news-blog/blog/co-creating-more-equitable-world-transformative-benefits-participatory-design (2021).

16. Demel, L. Citizen science: Building collective solutions. *UNDP Accelerator Labs* https://www.undp.org/acceleratorlabs/blog/citizen-science-building-collective-solutions (2021).

17. Alam, T. & Porras, J. Architecting and designing sustainable smart city services in a living lab environment. *Technologies* **6**, 99 (2018).

18. ABALOBI. ABALOBI – elevating small-scale fisheries through technology. *ABALOBI* https://abalobi.org/ (2021).

19. Castillo, N. M. & Vosloo, S. *Abalobi: Case Study by UNESCO-Pearson Initiative for Literacy*. https://unesdoc.unesco.org/ark:/48223/pf0000264639 (2018).

20. Community Tech. Community Tech. *Community Tech* https://www.communitytech.network (2024).

21. GenderMag. The GenderMag project. https://gendermag.org/gendermag.php.

22. CABI. *Module 7: How to Create a Data Inventory - Checklist*. https://www.datasharingtoolkit.org/wp-content/uploads/2021/03/CABI-Mod7-3-01-Checklist.pdf (2021).

23. Mitchell, J., Lesher, M. & Barberis, M. *Going Digital Toolkit Note: Measuring the Economic Value of Data*. (2022).

24. JoinData. JoinData. *JoinData* https://join-data.nl/en/homepage/ (2024).

25. Mulgan, G. & Straub, V. The new ecosystem of trust. *Nesta* https://www.nesta.org.uk/blog/new-ecosystem-trust/ (2019).

26. Google Cloud. What is data governance? *Google Cloud* https://cloud.google.com/learn/what-is-data-governance (2024).

27. European Parliament. GDPR data controllers and data processors. https://www.gdpreu.org/the-regulation/key-concepts/data-controllers-and-processors/ (2020).

28. European Parliament. Recital 32 - Conditions for Consent. *General Data Protection Regulation (GDPR)* https://gdpr-info.eu/recitals/no-32/ (2016).

29. Tait, J. Designing a data cooperative to help make homes more energy efficient. *Open Data Manchester* https://www.opendatamanchester.org.uk/designing-a-data-cooperative/ (2021).

30. Greenwood, D. (2022).

31. Co-operatives UK. The community shares handbook: asset lock provisions. https://www.uk.coop/resources/community-shares-handbook/2-society-legislation/24-asset-lock-provisions-cs.

32. Hunt, P. *Demutualisation and How to Stop It*. (2022).

33. Mondragon corporation. Mondragon corporation. https://www.mondragon-corporation.com/en/ (2024).

34. International Cooperative Alliance. *Cooperatives and Sustainability: An Investigation into the Relationship*. https://ica.coop/sites/default/files/2021-11/Sustainability%20Scan%202013%20Revised%20Sep%202015.pdf (2013).

35. Swash. Swash. https://swashapp.io/.

36. Dawande, M., Mehta, S. & Mu, L. Robin hood to the rescue: sustainable revenue-allocation schemes for data cooperatives. *SSRN Electron. J.* (2022) doi:10.2139/ssrn.4020556.

37. Kemppainen, L., Koivumäki, T., Pikkarainen, M. & Poikola, A. Emerging revenue models for personal data platform operators: when individuals are in control of their data. *J. Bus. Models* **6**, 79–105 (2018).

38. Nabben, K., Puspasari, N., Kelleher, M. & Sanjay, S. Grounding decentralised technologies in cooperative principles: what can 'decentralised autonomous organisations' (daos) and platform cooperatives learn from each other? in (The New School Platform Corporativism Consortium, Harvard Berkman Klein, 2021). doi:10.2139/ssrn.3979223.

39. Conger, K. A worker-owned cooperative tries to compete with uber and lyft. *The New York Times* (2021).

40. Parra-Moyano, J., Schmedders, K. & Pentland, A. Shared data: backbone of a new knowledge economy. in *Building the new economy* (2020).

41. MIDATA. Midata. *MIDATA* https://www.midata.coop/en/home/ (2024).

42. GOFAIR. FAIR Principles. *GO FAIR* https://www.go-fair.org/fair-principles/ (2016).

43. Our Streets Chorlton. Our streets Chorlton. https://ourstreetschorlton.co.uk/ (2024).

44. University College London. Anonymisation and pseudonymisation. *Data Protection* https://www.ucl.ac.uk/data-protection/guidance-staff-students-and-researchers/pra

ctical-data-protection-guidance-notices/anonymisation-and (2019).

45. European Parliament. Art. 4 GDPR – Definitions. *General Data Protection Regulation (GDPR)* https://gdpr-info.eu/art-4-gdpr/ (2016).

46. IBM. Obfuscating data method (Masking flow). https://www.ibm.com/docs/en/cloud-paks/cp-data/4.8.x?topic=options-obfuscating-data-method (2023).

47. Wen, T. Data aggregation. in *Encyclopedia of Big Data* (eds. Schintler, L. A. & McNeely, C. L.) 1–4 (Springer International Publishing, Cham, 2020). doi:10.1007/978-3-319-32001-4_296-1.

48. Salus.coop. Salus.coop. https://www.salus.coop/ (2024).

49. Morozov, E. Digital intermediation of everything: at the intersection of politics, technology and finance. in (Karlsruhe, Germany, 2017).

50. Mozilla Insights & Gahntz, M. Is that even legal? A guide for builders experimenting with data governance. *Mozilla Foundation* https://foundation.mozilla.org/en/research/library/is-that-even-legal/builders-guide/ (2023).

51. Mannan, M., Wong, J. & Bietti, E. Data cooperatives in Europe: a preliminary investigation. *Netw. Ind. Q.* **24**, 12–15 (2022).

52. Mannan, M. & Schneider, N. Exit to community: strategies for multi-stakeholder ownership in the platform economy. *Georget. Law Technol. Rev.* **5**, (2021).

53. Scholz, T. (2022).

54. Balkin, J. M. Information fiduciaries in the digital age. *Balkinization* https://balkin.blogspot.com/2014/03/information-fiduciaries-in-digital-age.html (2014).

55. Development Gateway. *Farmer Cooperatives in Uganda United in a Data Collaborative*. https://developmentgateway.org/wp-content/uploads/2023/02/Farmer-Cooperatives_CaseStudy_Feb2023.pdf (2023).