3

Data Capability Through Collaborative Data Action

In Chap. 2, we presented case studies of some of our data projects that involved working with non-profits and other types of organisations and re-using varied datasets. Each of these projects saw participants move from curiosity about data analytics, to a growth in confidence around using terminology, understanding techniques and having a grasp of non-profits' internal data resources. We argue that this represents the participants making progress in building aspects of the data capability of their organisations as well as understanding gaps. From our experience, successful results happen in data projects when people with diverse backgrounds and perspectives collaborate to explore issues of direct relevance to them, drawing on varied expertise, infrastructure and datasets. Organisations have existing data practices and resources, and so experimenting together with novel analytical techniques and types of datasets can help partners with a social mission to understand what to do next to extend and tailor their future data practices.

What we found through our projects with non-profits, then, is that *collaborative data action* supports the *building of data capability*. As depicted in our case studies, collaborations can draw across teams within a single organisation, across a set of like-minded organisation partners

J. Farmer et al., *Data for Social Good*, https://doi.org/10.1007/978-981-19-5554-9_3

and externally with researcher partners and others. In this chapter, we move from examples showing the sometimes messy business of non-profits working with novel datasets, to attempting to secure some concepts and processes that underpin non-profits working with data analytics. Thus, we explore here what we think data capability looks like for non-profits and provide our methodology for supporting capability to build through collaborative data action. In doing so, we suggest priority topics for non-profits to address, principally around establishing responsible data governance and being clear about ethics and consent.

Again, we note this is based on our practical work up to 2022, and from our base in Australia. Law and practices relevant to non-profit data analytics will be different in other countries and regions and are changing over time.

Understanding Data Capability

Drawing on our own research, we suggest that at an organisational level, *data capability* is a holistic resource. It involves having in place the interconnected aspects of appropriate *staff roles and skills, technologies*, and *data management practices and processes* to fulfil what an organisation needs and wants to do with data. In data science, *capability* has a dual meaning, relating both to human competencies and technical components like software, hardware and database systems. In our work, we retain this sense of data capability as multi-faceted and interconnected with multiple technical and human attributes. Data capability is additionally hard to pin down, we suggest, because it is situated or adaptive to context—that is, data capability will vary according with each non-profit's work, mission and vision in their operating context. We realise this can make data capability seem elusive and hard to measure, but we suggest it is most realistic to think of it as this combined, evolving, overall resource.

Data capability is related to data management and data governance. *Data management* is about having a system of internal practices and mechanisms for controlling data within an organisation. DAMA International describe centralised, distributed and hybrid models of data management, referring to the way parts of an organisation can work collectively and independently when managing and working with data (2017, p. 565). *Data governance* is the framework of ethics, safety and accountability practices that interweaves with and shapes how data management is done. We return to explore data governance as a foundation for data capability later in this chapter.

We suggest *data capability* is the *outcome* that non-profits should be aspiring to achieve as they increasingly use data analytics. However, it is not static, rather it is refreshed and continually reformed via *processes* of engaging with datasets and new ways of working with, and using, data. This means the data capability of an organisation formulates through adaptation and change via ongoing experimenting and learning with data. Considering our Chap. 2 case study projects as processes of learning, participants were generally more knowledgeable, confident and comfortable with using data and interpreting data analyses by the end of projects. While we did not have formal evaluation in all our projects, we witnessed instances of increased engagement with data among a wide range of staff members (not just data or IT professionals) and the adoption of more sophisticated data practices, often across teams and individuals who didn't normally work together. Participants developed agility and confidence in their ability to determine when and which types of data analytics and visualisations would be useful (or not) in specific contexts. They were generally more excited and animated about the potential of working with data into the future. Underpinning these findings, participants also talked about changes that would need to be made, particularly to their data management and data governance practices. Examples of this include questioning risk aversion in sharing datasets and talking about the need for strategic consideration of reconfiguring data governance. These are all aspects indicating the way data capability forms and provide examples of the multiple and small steps by which data capability develops in relation to context.

In our projects, we saw non-profits' data capability influenced through processes of practising with using their *own* internal datasets for insights about *their* problems and challenges. This seemed impactful, compared with participating in generic training modules or engaging with generic resource kits (as we tried in Case 2 described in Chap. 2). While building data capability still implies financial investment in technologies, infrastructures and skilled people, collaborative practice can help participants work out what their organisation needs and target their spending on priorities. Depending on who is involved in collaborative projects, progress in data capability can be activated strategically (from the top down) where senior managers participate, or from the ground up, through the action of practitioners in consumer and client-facing roles.

Responding to sectoral interest in increasing data analytics expertise across the non-profit sector, several frameworks have emerged for measuring and monitoring development of organisational resources related to having data capability (for example, see the work of https://data.org in the US). Some stakeholders—such as philanthropic foundations or non-profit representative bodies—seek to benchmark how individual non-profits compare in their *data maturity* against others in the sector. They also apply frameworks to identify sectoral strengths and gaps. Some assessment tools have rating scales, for example, with a low score for initial or ad hoc practices, to a higher score for systematically *managed* or *optimised* data practices (see, e.g., DAMA International's rating scale [DAMA International, 2017, p. 531]). In the UK, Data Orchard's *Framework for Measuring Data Maturity* in non-profit organisations (Data Orchard, 2019) aims for expert-level resources and practices or *mastery* as the goal, with maturity examined on dimensions including data uses, analysis, leadership, culture, tools and skills. We explored the difference we see between data capability and data maturity or data literacy in Chap. 1, saying why we prefer the idea of data capability as a goal for non-profits. This is mainly because we do not think data resources like human skills, technologies and practices should be fixed, but rather adaptive relative to each non-profit's context, strategy, mission, size and so on.

While we express reservations with static frameworks, one of our own collaborative research projects driven by perspectives from multiple Australian non-profits led to the creation of a broad data capability framework (Yao et al., 2021). This identifies attributes participating non-profits considered central to their data work. These are assigned to four domains: (1) *access* to quality data; (2) data *skills* and ability; (3) effective *technology* systems, tools and data infrastructure; and (4) responsible data *governance* (see Yao et al., 2021). However, even given this framework, we have found more generally in our work with non-profits that rather than embracing levels of attainment on a fixed scale, many emphasise they have nuanced and varying needs and goals for data use. Consequently, the value of frameworks, for them, was suggested as offering shorthand checklists against which to reflect on organisational strengths and gaps against an indicative industry standard.

Building the more holistic resource of data capability also enables non-profits to influence and activate beyond their own operational matters. For larger organisations, this could involve sharing data expertise with other, smaller organisations and helping to develop sector-wide collective responses to social problems. Alternatively, it could involve developing shared data resources or data collaboratives like the *Humanitarian Data Exchange (HDX)* (https://data.humdata.org/). Having data capability provides a foundation for a non-profit to partner with their clients and communities on data projects with wide social benefit. Hendey et al. (2020) depict this as non-profits contributing to a wider social mission of enabling *community data capability*. While no single model of community data capability exists, the authors argue that when data capability and resources are democratised and available to those who can benefit, "communities will be better equipped to partner with foundations, apply data to understand issues, and take the actions needed to achieve the ambitious outcomes that [philanthropic] foundations seek" (Hendey et al., 2020, p. 1). Non-profits are well placed, due to their work and missions, to drive community data capability goals.

A Collaborative Data Action Methodology

Our case studies in Chap. 2 show where we have worked in collaborations with non-profits, sometimes with staff members across teams of one organisation and sometimes across organisations. In those projects, we observed teams and groups addressing a data challenge, but also in the process, developing or at least influencing their data capability. Some of the impacts of working collaboratively are highlighted at the end of Chap. 2. Observing the projects, their direct outcomes and wider impressive impacts has made us committed to collaborative working; and in this section, we talk specifically about our collaborative data action methodology.

There could be a range of different ways that non-profits could gain data capability through collaborative working. This could be through working with other non-profits with large or specialist data science teams, working more effectively across teams within their own organisations, or accessing data collaboratives or external *data for social good* initiatives (see this book's appendix). The point is to engage with others with a shared social mission and to gather a team of people that combines useful knowledge, skills and perspectives.

There are some very practical implications of collaborating that we have already alluded to. These include accessing others' expertise and resources to help improve your own organisation's access to costly resources and to learn what you need by efficient contextualised learning. There are also wider benefits of collaborating. Firstly, the field of data analytics is moving so fast at present that it requires dedicated specialists to keep up. This is just data science, of course, and the fields of social justice and addressing a social mission have also changed dramatically in response to the pandemic and its ongoing effects. A simple benefit of collaborating is that it gives access to a wider range of human resources to keep up with changes in knowledge and techniques across fields of expertise and practice. Collaborating is also a way to help keep small, potentially niche non-profits operating as the sector becomes more corporate and favours larger organisations. Finally, and importantly, organisations collaborating with data for social good help to build the field. Working together generates new networks, social capital and communities of practice between organisations that will impact more widely to foster community data capability.

In our projects, we use a process of collective 'learning by doing' or *collaborative data action*. The process allows for experimentation and adaptation. It allows individuals within non-profits, including senior managers and board members, to see how working with data can help to integrate their operations and services across departments (i.e., wider benefits). And it can help to empower and activate grass-roots practitioners in incorporating data work as part of their daily practice.

While data projects will vary in their precise process due to different participants, questions, data and timelines, we have found there are a consistent set of main activities that punctuate collaborative data action in our data projects with non-profits. Figure 3.1 outlines these main activities, giving an approximate chronology.

At this point, we highlight that we have mainly used the collaborative data action methodology when working with organisations seeking to find out whether data analytics is useful for them. This could suggest it works best for those setting out from *a low base*; however, that is not the whole story. For example, the bank in Case Study 3 had a large and sophisticated data analytics team, and in Case Study 1, we worked with the business insights unit of government, a team specialised in data analytics to inform policy. Rather, then, perhaps the collaborative data action methodology is best regarded as a mechanism for experimenting with data analytics. Experimenting can involve starting out, but it can also involve trialling different techniques for data analysis or addressing

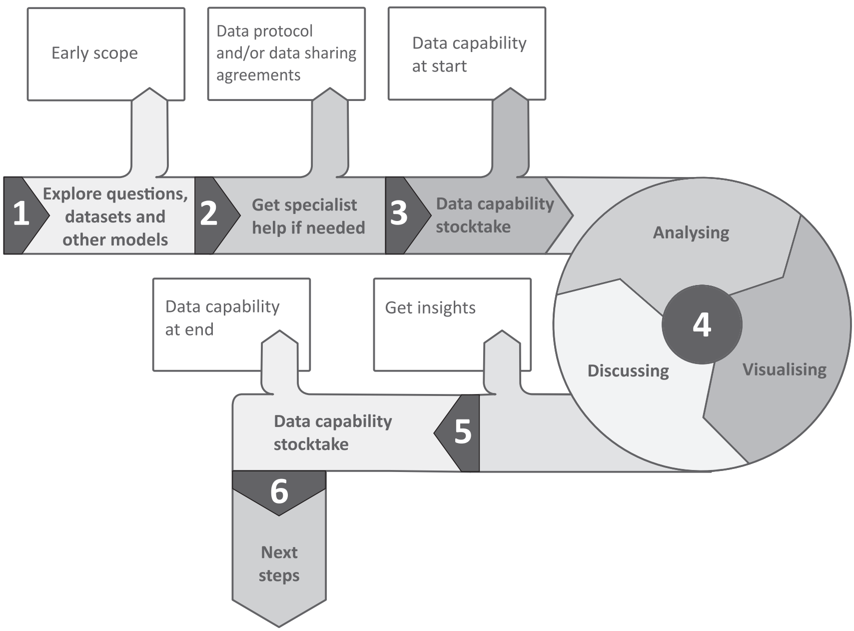

Fig. 3.1 Process of collaborative data action for non-profits' data projects

more ambitious goals. Thus, collaborative data action can involve organisations that are skilled-up and advanced in working with data. Of course, a key element here is that an organisation can access a range of knowledge, technology or other resources that can help to work with data in different ways or inject other types of knowledge (e.g., from social science or community practice) into data analytics.

In our projects, we tried out various activities as part of processes of experimenting and collaborating in data projects. Some approaches we initially included turned out to be blind alleys—for example, the general educational webinars we provided in Case Study 2 turned out to be less well-received than learning by doing experienced with participants in addressing their organisations' challenges and using their data. Ultimately, we arrived at a methodology comprising a relatively consistent set of activities that helped to produce project outputs and processes and within which participants said they experienced learning and enjoyment.

Steps in our collaborative data action methodology involve different kinds of actions (see Table 3.1). Some steps involve *exploring.* Step 1, for example, is about simultaneously exploring ideas from previous case studies, questions to focus on, and useful datasets all in order to test the feasibility of undertaking a data project and deciding its initial scope.

Step 2 involves turning to specialist experts examples, and precedent for help to formally get started. If a project is being undertaken internally and involves just one organisation, then a data protocol should be drawn up establishing what is to be done with data and why. If a project involves collaborating and sharing data across organisations, then data sharing agreements will be required that allow partners to work together with internal datasets. Data sharing is notoriously complex and requires engaging with legal principles influenced by the laws and guidance that apply in different geographical jurisdictions. Individual organisations will also have their own protocols and require compliance with sectoral guidance. We have indicated some current resources that can help to think about data sharing and what is required in data sharing agreements in the appendix. Data sharing across organisations is also revisited later in this chapter.

In our projects we also found that it was useful to build in some formal *stocktake* or evaluation 'before and after' opportunities to facilitate reflection at the start and end of data projects. This enables participants

Table 3.1 Steps in the process of collaborative data action for non-profits' data projects

Step	Actions	Goal/achievement
Early steps		
1. EXPLORE initial question or focus, potential data sources and similar data projects	Consider what topics or questions the data project might target and what internal and open datasets there might be that could address the question. Explore examples of other data projects and their output visualisations and engage with potential data collaborators with a shared interest and useful skills	Draft early scope of a project, including questions, datasets and collaborators across teams and/or other organisations
2. Bring in SPECIALIST HELP for establishing data protocols or agreements	Work with a legal team and data collaborators to establish data protocols and, if needed, data sharing agreements matching jurisdiction/sector legal requirements	Have agreed data protocol and/or data sharing agreements
3. Pre-project data capability STOCKTAKE	Conduct an early 'stocktake' to establish all participants' goals, data challenges and gaps in capability	Summary of data capability at the start
Doing the project		
4. ITERATE through cycles of analysing & visualising datasets, using DATA WALKS to EXPLORE and then analysing other datasets and/or ADAPTING visualisations and questions	Begin initial data analysis using identified data sources and generate visualisations to discuss findings as a group. Then repeat this process until a focused question has been addressed or insights gained, that is, until the group is sufficiently satisfied they have attained their goals in the data project	Identify insights and visualisations to address focus questions
End of project		
5. End of project data capability STOCKTAKE	Conduct follow-up stocktake to find out what has changed, any learning and remaining gaps	Summary of changes in data capability
6. NEXT STEPS	Think about what has been learned and what should be done next	Acknowledge outcomes of the data project and agree next steps

to identify changes in their attitudes and practices at individual and organisational levels. This stocktake can be simple and involve thinking about and documenting concerns about data, aspirations for using data and assessments of expertise and readiness. At the end of projects, it can be about what was learned and what remain gaps. Stocktakes are at steps 3 and 5 of our methodology. We did not include formal data gathering stocktakes in our early projects (e.g., Case Study 1), but we discovered its value in Case Study 2 and then applied this learning in Case Study 3 and other projects since.

Step 4 involves *iteration* of several activities of working with datasets, aiming to answer questions and point to next steps. It involves analysing and visualising data and then exploring and discussing results. Once analyses and visualisations have been explored, it is usually necessary to cycle back a few times to identify other useful datasets and analyse and visualise these—all with the target of getting closer to an 'answer' to questions set or topics to be explored via the data analyses and to find out more about the topic(s) involved in exploring a question.

In our projects we employed cycles of workshops using an approach inspired by the data walks method of the Urban Institute's National Neighborhood Indicators Partnerships (Murray et al., 2015). Data walks involve workshop discussion where participants are shown visualised analyses, and encouraged to ask questions, engage with what *they see* in the data and sense-check this given their grass-roots knowledge. Iterative rounds of data analysis followed by discussion help participants to make sense of data that has been analysed and visualised and to discuss with each other, the stories they perceive to be told in the data. Visualisations are an important part of data walks, as diagrams, geospatial maps and graphs tend to be commonly accessible to participants from different backgrounds. In our projects, data walks were useful for considering topic-based insights but also for stimulating technical queries about datasets and exploring issues about data collection affecting interpretation of analyses.

Based on feedback on analysed and visualised data from the workshops, new datasets may be identified and analysed, new types of analysis might be conducted with the same datasets or different visualisation techniques might be employed. Then new analyses and visualisations would be brought back for further discussion and sense-making at a workshop, with the idea being to cycle through multiple workshops until a question or focus topic has been sufficiently addressed. Open-ended cycles of iteration can be challenging to explain in funding applications and contracts, so it may be useful to consider that in our projects we found three to four iterative cycles generally produced useful findings. After more than three to four cycles, the project might lose impetus and participants might lose interest.

Exploring questions and datasets collaboratively in workshops helps to generate a shared understanding and language around data use and outcomes sought. The collaborative methodology ensures that each participant shares their perspective in these sessions and their take on featured questions and data. This means that no single department within an organisation or dominant partner, if working across organisations, imposes their viewpoint. Taking an exploratory approach can generate wider buy-in by showing that different participants can have different, equally valid, ways of understanding a question, problem or challenge being addressed. Understanding can be gained here about how problems are multi-faceted, prompted by discussing insights suggested by data analyses.

This working between question(s) and dataset(s) that we describe involves processes of *adaptation*, with a goal of matching data with questions. Sometimes the adaptive process leads to framing a question in a different way. At other times, there is a realisation that a whole and perfect dataset to answer a pre-defined question does not exist, prompting a turn to other data that can *inform* about a question if not answer it directly. An example here was where the state government participants in Case Study 1 came to realise that a comprehensive dataset precisely aligning with changed attitudes to family violence did not exist. Instead, we harnessed Twitter data and news media data with textual data analytics to show a quite granular change in topics discussed over time. At the same time, we know there are caveats about some of these datasets. For example, Twitter users are a self-selecting, more policy-aware community. The government itself periodically conducts a Community Attitudes Survey covering attitudes to family violence but, again, responses in that dataset are from self-selected participants who tend to be older and more educated. Together, the data from the three sources (Twitter, news media, community survey) can be *triangulated* to give richer, though still not comprehensive, information about the extent of discussion (in this case related to family violence), variety of topics discussed and responses to different types of policy and other events.

The adaptive way of working between topics and questions that we adopt is one way that our approach is potentially distinct. Other data project methodologies we have seen emphasise pursuing and identifying *a precise problem or question* before proceeding to data analysis (e.g., The GovLab, n.d.). While it is important to have a broad initial focus, we have found it can be difficult for non-profit partners to identify specific questions or *pain points* at the start of a data project. This can be because participants don't have a grasp of what data might be available, what might be possible (and not possible) with data analytics and may need time to understand the work of other participants. In our experience, focus for projects does happen, but it emerges or sharpens through working with data and discussing questions iteratively and learning what is possible and useful. Being open as to focus can be challenging for non-profits to justify in funding applications, so a useful strategy is to identify a broad topic to explore from the start.

Following the end of project stocktake at step 5, the conclusion of the process is to acknowledge what has been achieved in terms of data product outputs and wider outcomes in relation to learning or partnerships and to decide what next steps are appropriate, if any.

Finding Your Data Collaborators

In this book, we propose that building data capability should not be a solo practice. Building data capability could be done through working on experimental data projects and these might benefit, depending on their scope and goals, from the skills and perspectives of a range of different people, teams and organisations. Preferably, this would also include lived experience consumers, clients and citizens because they will help to make more insightful, ethical data products and extend data capability within the community. In Chap. 2, we showed that the collaborations we have worked within took multiple forms. They involved working across departments *inside* an organisation (as with Good Cycles and Yooralla in

Case Study 2, and multiple departments and agencies of government in Case Study 1) and working *across* non-profits and other community organisations (as in the City of Greater Bendigo data collaborative project in Case Study 3). In each case, our university-based social data analytics team brought expertise in data science and social science, as well as access to technologies and safe, secure practices. The collaborating partners brought their expertise which also involved data analytics skills and understanding of problems and contexts. When we were re-using non-profits' internal datasets, their staff could inform about how data was collected and what was included and excluded in datasets.

We term the various participants—people, teams, organisations—in data projects as data collaborators. While a range of perspectives makes the collaboration more than the sum of its parts, clearly the main thing we are focused on is the potential offered by injecting advanced know-how about data science and analytics. It is a premise of this book that the projects we describe are about building (greater) data capability for non-profits. In our projects, the university team brought access to advanced data science knowledge, technology and practices. While here we mainly focus on university teams, there is a range of ways to access collaborating partners with data science expertise. Non-profits might partner with other, perhaps larger, non-profits that have specialist data analytics teams or collaborate together to approach some external entity with expertise. In the appendix, we suggest some data analytics initiatives that have a particular mission to build data analytics capability of the non-profit sector. Initiatives working to support data capability development are sometimes termed *data intermediaries* or *data institutions* (Hardinges & Keller, 2022). These might offer opportunities for mentoring and learning in partnerships (Perkmann & Schildt, 2014; Susha et al., 2017), although some data intermediaries are more engaged as *brokers* between organisations and data owners (Sangwan, 2021). In encouraging collaborations between non-profits and other social sector actors to grow data capability and community data capability, we align with the concept of the organisational partners envisaged in the National Neighborhood Indicators Partnerships. Many of those partnerships combine local community organisations, non-profits and councils working with university social data analytics labs (Arena & Hendey, 2019).

As university researchers ourselves, we recognise and suggest the potential of seeking out a university social data analytics lab to work with. The opportunity is that such labs will often share the social mission orientation of non-profits, and there are many examples of labs situated in universities around the world. Some university data analytics labs will be actively looking to partner for access to 'real-life' projects for training data science students. As one example, the Center for Urban and Regional Affairs (CURA) at the University of Minnesota (https://www.cura.umn.edu) links academics and students with community organisations to generate data analytics projects, specialising in data for neighbourhood planning. Other examples of university data labs working with non-profits can be found in the literature; for example, Tripp et al. (2020) describe a partnership between an education and literacy non-profit and the West Georgia University's Data and Visualisation Lab. Of course, generally universities do still require funding to work on data projects. This could come directly from a non-profit or partnerships could be formed with university labs to apply, together, for funding.

Different partners collaborating with data and sharing knowledge and skills generates new *boundary spaces* (Susha et al., 2017). These enable novel combined skillsets to emerge, helping to grow a future workforce of people that understand both non-profit work and data analytics. Research literature describing *how to do* data analytics for social good emphasises the significance of a diverse team, including data scientists, social scientists, practitioners and lived experience consumers and clients (e.g., Williams, 2020).

Responsible Data Governance

In the last part of this chapter, we focus on practices that all non-profits will already have considered in some way if they are working with data: these are practices of data governance. Data governance is understood here as having the systems and processes so that an organisation can ensure data is managed and analysed responsibly, legally and ethically. It involves having clear mechanisms through which an organisation, and its people, are held to account about the production and use of data. We focus on data governance here because it is a priority consideration for an organisation working to re-use its data. Having appropriate data governance in place is a necessary precursor to working in data projects, particularly when engaging with other organisations in a collaboration. It is also a feature that organisations can start working on without having to wait to find data collaborators to work with.

Having responsible data governance enables an organisation to have safe and secure data, accountability, quality assurance and ethical data practice. Active engagement across organisations in data governance will result in a positive data culture, with all staff, clients, consumers, managers and board members engaged in well-considered, ethical data work.

Co-ordinated practices of responsible data governance should be thought through and implemented by any organisation collecting and using data. Data governance sits around, permeates and directs data management, including affecting who works with data (roles and skills), technologies and how they are used, and the nature of practices and processes in handling, storing and analysing data. Governance will need to be able to respond to changing organisation requirements to use different datasets with different types of analyses. Data governance needs to be integral to organisational governance, not seen as separate, as it relates to whole of organisation best practice and accountability. With increased production, storage and use of data, and the consequent potential for many forms of data harm, data governance has become an important aspect of organisational governance (Redden et al., 2020). This includes aligning and interweaving data practices with the protocols and policies that guide an organisation's practices around ethics, risk management, compliance, administration and privacy (Governance Institute of Australia, 2022).

The significance of data governance makes it a strategic organisational issue, and the priority data governance is given by organisations will determine what they can do with data. The values inherent in how data governance is implemented shapes the goals and outcomes of using data. This includes ways of viewing relationships—customers and clients can be 'mined', and their data 'extracted', or they can be consenting collaborators, with their needs aligned to how data is used.

Depictions of data governance in the research literature can suggest a commercial emphasis inappropriate for the non-profit sector. For example, Otto (2011, p. 47) defines data governance as "a companywide framework for assigning decision-related rights and duties in order to be able to adequately handle data as a company asset" (cited in Alhassan et al., 2018, p. 301). Objectifying data in this way, as a kind of commodity, serves to disregard the integrative relationship between data, people and services. It might be said, therefore, that non-profit data governance models compare, but also differ, in ways from those of commercial organisations, with differences driven by mission, context and vision of each non-profit.

While frameworks for data governance tend to be internally focused, the requirement for formal policies and protocols is increasingly driven by interactions with the external environment. This is especially true in relation to embarking on data collaborations involving other organisations and sharing datasets (Verhulst, 2021). Indeed, increasingly, experts advocate for data stewards as a kind of data governance role for organisations serious about developing data capability (Verhulst et al., 2020). "Data stewardship is a concept with deep roots in the science and practice of data collection, sharing, and analysis. Reflecting the values of fair information practice, data stewardship denotes an approach to the management of data, particularly data that can identify individuals" (Rosenbaum, 2010, p. 1442). Data stewards would be responsible for understanding the datasets that exist in organisations and ensuring their quality. One role for organisational data stewards would be in bringing internal datasets into collaborations across organisations to facilitate data collaboratives and data sharing.

While designating a data steward signifies organisational acknowledgement that data governance is important and demands an owner, the holistic nature of data governance suggests it as also collective action issue. As touched on in Chap. 1, clients, customers and other people in the data of non-profits and involved in its collection, should be included in designing data governance that assures fairness and empowerment. Some researchers have demonstrated "the value in theorizing data governance as a collective action problem and argue for the necessity of ensuring researchers and practitioners achieve a common understanding of the inherent challenges, as a first step towards developing data governance solutions that are viable in practice" (Benfeldt et al., 2020, p. 299).

Topics at the heart of responsible data governance are ethics and consent and are featured below. Clarity about ethics and relationships of consent and trust is essential because of the imperative of accountability to all of the people who are stakeholders in the data. Getting ethics and consent right sets non-profits up to achieve in more ambitious, innovative and strategic efforts of working with data beyond basic use of internal datasets—that is, looking to data collaboratives and data sharing.

Data culture is closely related to data governance. When data governance is working well, it becomes embedded and part of the everyday practice of organisations, contributing to a positive data culture. Clearly data culture can be of varying quality, dependent on attributes such as inclusion in governance, ethics-orientation and embeddedness in roles, operations and strategy. We understand data culture here as the organisationally embedded ways of understanding and working with data ethically and safely. Central to having a positive data culture is instilling and embedding genuine concern about the relationship between the people who generate the data (bearing in mind Williams' assertion that "data are people" [Williams, 2020, p. 220]) and what can thus be done with data. Disciplined thinking about consent and trust must be established and maintained. Data culture relates to the values of organisations around enabling and empowering people (staff, clients, customers and others) and accountability to these stakeholders. While we found little written about organisational data culture and its development, it seems an issue that is close to consideration of organisational ethics.

Data Ethics and Consent

Issues of ethics and consent are fundamental to consider from the start of any data collection. They are difficult to 'retrofit' if a non-profit decides it wants to re-use data originally collected to measure outputs or for statutory reporting. Clearly as well, addressing these issues is not about organising so that a non-profit can have the data it wants to work with. The question of who owns the data, and is *in* the data, is the ethical issue here. As highlighted in Chap. 1, work is ongoing internationally to partner with people who are (in) data to drive its ethical collection and use. Indigenous scholars have perhaps gone furthest in showing why and how marginalised groups should be driving collection and use of data about them. For example, Kukutai and Taylor (2016) documented the importance of affirming Indigenous people's rights to self-determination via recognition of data sovereignty.

Some practical guidance and resources to help non-profits achieve ethical data use and re-use have been developed by data initiatives internationally (e.g., National Neighborhood Indicators Partnership, 2018; NESTA, 2022; and see the appendix). In our own work in collaborations with non-profits, we have found that some materials about ethics and consent can be high-level, too general or too specific in their nature for application across diverse contexts. As a body of advice, the sheer amount of guidance can even seem overwhelming. Perhaps because of this, among the communities of data practice where we have participated, non-profits tend to share and adapt data management, privacy and security policies among their networks and to develop norms around data collection and use through cumulative processes. Data ethics is not always explicitly discussed, even if care and responsibility is taken in all data practices. Here, we suggest how to begin to think about and apply data ethics, irrespective of precise frameworks or protocols, by focusing on establishing relationships of care and consent in data production and use.

Firstly, there are legal considerations in using personal data and data governance is entwined with regulation and increasingly the subject of law reform across different global jurisdictions. Laws governing personal data have dealt mainly with issues of privacy and cybersecurity but are becoming more complicated as technology develops and services become 'digital-first'. Because these are jurisdiction-specific, all we can suggest here is to consult jurisdictional sector representative bodies and the government agencies established to guide and inform adherence to relevant laws. If working with *sensitive data*—for example, personal data, especially where it concerns health, race, sexuality, beliefs and associations—data ethics and data management practices (like secure or encrypted storage, de-identification and access protocols) are high priority. Non-profits should consider working with a legal advisor with relevant understanding of data, information and privacy regulation.

Beyond compliance with relevant data regulation, there is growing recognition of the need to begin with ethical frameworks and develop policies and practices for data use that involve carefully established trust and consent. By consent we do not simply mean the kinds of contractual agreement documents or pages that people sign or click 'OK' to engage with a service. These are instruments for establishing consent, but we are referring more broadly to the relationships developed within an organisation and with customers, clients and citizens around data collection and use.

Gaining consent for data use is a *process* for ensuring good data practices and relationships. It does not happen just once but is maintained and re-established as part of managing client and customer relationships and ensuring informed agreement with any new use of data. This is often approached through the establishment of norms (based on an organisations' values) of what an organisation *should* do to work safely with personal data, and with *care*. Two useful guiding principles are that any data collected should be necessary, and the purpose should be transparent and communicated clearly to those involved in generating the data or to whom it refers. This requires deciding what data is to be collected and its purpose, and an organisation may have detailed policy documents and ethical frameworks to help guide those decisions. As raised in Chap. 1, non-profits should be working towards involving consumers or clients (i.e., often the subjects in and of non-profits internal datasets), in codesigning these practices, avoiding tokenistic forms of inclusion.

As part of data governance, a comprehensive set of data ethics protocols and policies can help to drive a positive organisational data culture. With data collection increasing, data ethics scholars have identified core concerns to be addressed. Mittelstadt and Floridi (2016) emphasise informed consent, privacy (including data anonymisation and data protection), ownership and control over data, epistemology and objectivity (or data quality), and data-driven inequality "between those who have or lack the necessary resources to analyse increasingly large datasets" (Mittelstadt & Floridi, 2016, p. 303). Franzke et al. (2021) describe the development of a Data Ethics Decision Aid (DEDA), used to reflect on and guide decisions about data projects in the governmental context. The Open Data Institute's (2019) Data Ethics Canvas identifies 14 categories to help assess ethical aspects of using data in an organisational or government context.

There are increasing moves for organisations to collaborate to share reused data generated through their work. Our City of Greater Bendigo data collaborative (see Case Study 3 in Chap. 2), for example, was developed because seven community organisations wanted to find out whether pooling their data could help to generate new insights about community resilience. There are important ethical dimensions to such data re-use in the context of data sharing. There are logistical aspects to data sharing—why do it, what data and for what kinds of analysis? But data sharing and re-use are underpinned by governance and ethical issues first, because data use is contingent on the arrangements in place to ensure data is treated ethically, safely and with care. Foremost is clarity about whether consent for different types of use has been established or needs to be (re-)established with those who are the subjects of the data. Consent might have been established for a primary purpose but not for a secondary purpose. In Europe, the General Data Protection Regulation (GDPR) laws restrict data re-use and suggest re-establishing consent for secondary use (European Parliament and the Council of the European Union, 2016). In that jurisdiction, data can be re-used for a secondary purpose if its use relates to the primary purpose and a person would reasonably expect it to be used for the secondary purpose. For health information or other sensitive information, re-use is contingent on a direct link with the primary purpose for data collection.

Ensuring that ethics and consent issues are well considered, clear and codified, and comply with jurisdictional data legislation and practice is significant to guiding a non-profit's internal use of data. This becomes crucial when starting to work with other organisations to re-use data in collaborations. Ethics and consent practice govern the extent to which analyses of a non-profit's internal data can be undertaken, shown or shared with other organisations. While this might sound straightforward, consider what is potentially hidden in that deceptively simple idea of showing or sharing. In our City of Greater Bendigo Case Study 3 (see Chap. 2), it was one thing to look at each organisations' visualised data analyses in a workshop of seven organisations' representatives, but we then had to work out whether the visualisations could be seen by other staff or even explored in wider community engagement exercises. If visualised analyses of data could be shared, then in what formats? For example, ultimately percentages at suburb level were converted into an index of high to low relative quantities (e.g., in relation to wealth or demand for types of services) in our visualisations. This meant these could be shared beyond immediate workshop participants. This decision was taken on the basis of adhering to consents given/obtained for each dataset. The decision also responded to perceived potential reputational risks where community members might react adversely to seeing visualisations of datasets, for example, bank or service demand data, even if completely unidentifiable to individuals or households.

Data Sharing for Collective Gain

Given the issues just raised about data sharing in the example of Case Study 3, finally in this chapter we focus specifically on the data governance issue of consent and secondary use of datasets and data sharing. Because an organisation might want to move beyond re-using their own internal data and collaborate with others around data, obtaining appropriate consent is fundamental to data collection. A broad framework of thinking that we have used to guide our projects is the *Five Safes* model, initially developed by the UK Data Service (2017) to enable researchers to access government and sensitive data. This model was later adopted by the Australian Office of the National Data Commissioner as principles for access to and re-use of public sector data while maintaining data privacy and security. Though developed for public data sharing, the principles of the Five Safes are equally applicable as a guide to safe data sharing in the non-profit sector. It helps as a high-level framework to evaluate major risk areas and to identify steps to minimise the risk of data re-use. The Five Safes model draws attention to issues of sharing data in the domains of:

- Projects: ensuring data is shared for an appropriate purpose that delivers a public benefit.
- People: ensuring those using the data have the appropriate authority to access it.
- Settings: ensuring the environment in which the data is shared minimises the risk of unauthorised use or disclosure.
- Data: ensuring appropriate and proportionate protections are applied to the data.
- Output: ensuring output from the data-sharing arrangement is appropriately safeguarded before any further sharing or release.

Data collaboratives have become more widely discussed, as organisations recognise the value of working together to address community challenges. In our case studies, we showed an example of a community data collaborative where a range of organisations united around their internal datasets to explore for insights about community resilience. Our data collaborative projects use our Data Co-op platform (https://datacoop.com.au) that has software, hardware, management practices, multidisciplinary skills and data governance to support safe data sharing. Funded to the tune of over AU$1,000,000 by the Australian Research Council and five universities, this scale of investment in data collaborative infrastructure is outside the scope of most non-profits. We propose this supports our suggestions above that non-profits seeking to develop more ambitious data analytics projects could usefully collaborate to achieve more ambitious and complex projects.

Data collaborations can have various forms and work together for different reasons (Susha et al., 2017). Verhulst and Sangokoya (2015) give an example of humanitarian organisations working to share data for disaster relief. NCEL, Nepal's largest mobile operator, shared anonymised mobile phone data with the non-profit Swedish organisation Flowminder. With this data, Flowminder mapped where and how people moved in the wake of the disaster and shared this information with the government and UN agencies to assist their relief efforts. The Data Collaborative between NCEL and Flowminder allowed humanitarian organisations to better target aid to affected communities—saving many lives. While there is great potential and promise for data sharing, Verhulst (2021) highlighted that collaborating with data is one of the main challenges that (big) data initiatives for public good currently face.

As part of the appendix, we highlight some examples of resources and tools about data sharing that could be used by non-profits to find more information and examples, including example data sharing agreements.

Key Takeaways from This Chapter

In this chapter, we aimed to move beyond a rationale for non-profits getting involved in data analytics (Chap. 1) and illustrating how this can be done (Chap. 2). We explored data capability, a collaborative data action methodology, data governance, ethics and consent. The key points to take away from this chapter are presented below.

Key Takeaways

- Data capability for non-profits is a holistic resource that involves interconnected aspects of appropriate staff roles and skills, technologies and data management practices and processes that match needs, mission and strategy. It isn't static because it changes in relation to context, work and goals.
- Collaborating in data projects (collaborative data action) is a way to build data capability and to learn what is needed to achieve data capability. It is useful because it targets real challenges of participating organisations or departments and brings together varied expertise and different perspectives on challenges.
- Putting in place a sound data governance system is vital for managing data responsibly, legally and ethically and underpins a shared organisational data culture. More than a set of processes, it involves strategic thinking about relationships between a non-profit and its consumers, clients, customers and communities.
- Laws governing consent and access to data in jurisdictions are significant to working ethically. Alongside this, formulating consent and data sharing processes ideally involves co-design, including with people represented in the data.

The next and last chapter reflects on overall learnings, gives practical advice about starting or proceeding, and looks to the future and its challenges and possibilities.

References

Alhassan, I., Sammon, D., & Daly, M. (2018). Data governance activities: A comparison between scientific and practice-oriented literature. *Journal of Enterprise Information Management, 31*(2), 300–316. https://doi.org/10.1108/JEIM-01-2017-0007

Arena, O., & Hendey, L. (2019). *A look at the diversity of NNIP*. National Neighborhood Indicators Partnership, Urban Institute. Retrieved April 14, 2022, from https://www.neighborhoodindicators.org/sites/default/files/publications/A%20Look%20at%20the%20Diversity%20of%20NNIP_FINAL.pdf

Benfeldt, O., Persson, J. S., & Madsen, S. (2020). Data governance as a collective action problem. *Information Systems Frontiers, 22*(2), 299–313. https://doi.org/10.1007/s10796-019-09923-z

DAMA International. (2017). *DAMA-DMBOK: Data management body of knowledge*. Technics Publications.

Data Orchard. (2019). *Data maturity framework for the not-for-profit sector* (Version 2). Retrieved April 14, 2022, from https://www.dataorchard.org.uk/resources/data-maturity-framework

European Parliament, & the Council of the European Union. (2016). REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union, L199/1–L119/88. Retrieved April 14, 2022, from https://eur-lex.europa.eu/eli/reg/2016/679/oj

Franzke, A. S., Muis, I., & Schäfer, M. T. (2021). Data Ethics Decision Aid (DEDA): A dialogical framework for ethical inquiry of AI and data projects in the Netherlands. *Ethics and Information Technology, 23*, 551–567. https://doi.org/10.1007/s10676-020-09577-5

Governance Institute of Australia. (2022). *What is governance?* Retrieved January 19, 2022, from https://www.governanceinstitute.com.au/resources/what-is-governance/

Hardinges, J., & Keller, J. R. (2022). *What are data institutions and why are they important?* The Open Data Institute. Retrieved March 29, 2022, from https://theodi.org/article/what-are-data-institutions-and-why-are-they-important/#:~:text=Data%20institutions%20are%20organisations%20that,into%20our%20theory%20of%20change

Hendey, L., Pettit, K. L. S., Cowan, J., & Gaddy, M. (2020). *Investing in data capacity for community change*. Urban Institute. Retrieved April 14, 2022, from https://www.urban.org/sites/default/files/publication/102347/investing-in-data-capacity-for-community-change_1_1.pdf

Kukutai, T., & Taylor, J. (Eds.). (2016). *Indigenous data sovereignty: Toward an agenda*. ANU Press.

Mittelstadt, B. D., & Floridi, L. (2016). The ethics of big data: Current and foreseeable issues in biomedical contexts. *Science and Engineering Ethics, 22*(2), 303–341. https://doi.org/10.1007/s11948-015-9652-2

Murray, B., Falkenberger, E., & Saxena, P. (2015). *Data walks: An innovative way to share data with communities*. Urban Institute. Retrieved April 14,

2022, from https://www.urban.org/research/publication/data-walks-innovative-way-share-data-communities
National Neighborhood Indicators Partnership. (2018). *NNIP lessons on local data sharing*. Retrieved April 14, 2022, from https://www.neighborhoodindicators.org/library/guides/nnip-lessons-local-data-sharing
NESTA. (2022). *Data analytics*. Retrieved August 5, 2022, from https://www.nesta.org.uk/project/data-analytics/
Open Data Institute. (2019). *Data ethics canvas*. Retrieved January 18, 2022, from https://www.theodi.org/wp-content/uploads/2019/07/ODI-Data-Ethics-Canvas-2019-05.pdf
Otto, B. (2011). Organizing data governance: Findings from the telecommunications industry and consequences for large service providers. *Communications of the Association for Information Systems, 29*, 3. https://doi.org/10.17705/1CAIS.02903
Perkmann, M., & Schildt, H. (2014). Open data partnerships between firms and universities: The role of boundary organizations. *Research Policy, 44*(5), 1133–1143. https://doi.org/10.1016/j.respol.2014.12.006
Redden, J., Brand, J., & Terzieva, V. (2020). *Data Harm Record (Updated).* Retrieved January 19, 2022, from https://datajusticelab.org/data-harm-record/
Rosenbaum, S. (2010). Data governance and stewardship: Designing data stewardship entities and advancing data access. *Health Services Research, 45*(5p2), 1442–1455. https://doi.org/10.1111/j.1475-6773.2010.01140.x
Sangwan, S. (2021, April 27). How to know you are a 'data intermediary' under the Data Governance Act. *The Privacy Advisor*. https://iapp.org/news/a/how-to-know-you-are-a-data-intermediary-under-the-data-governance-act/
Susha, I., Janssen, M., & Verhulst, S. (2017). Data collaboratives as a new frontier of cross-sector partnerships in the age of open data: Taxonomy development. *Proceedings of the 50th Hawaii International Conference on System Sciences* 2017*,* Waikoloa Village, Hawaii, United States. https://doi.org/10.24251/HICSS.2017.325
The GovLab. (n.d.). *Phase 1: Demand*. Retrieved April 14, 2022, from https://datacollaboratives.org/canvas.html
Tripp, W., Gage, D., & Williams, H. (2020). Addressing the data analytics gap: A community university partnership to enhance analytics capabilities in the non-profit sector. *Collaborations: A Journal of Community-Based Research and Practice, 3*(1), 11. https://doi.org/10.33596/coll.58
UK Data Service. (2017). *What is the Five Safes frameworks?* Retrieved January 18, 2022, from https://ukdataservice.ac.uk/help/secure-lab/what-is-the-five-safes-framework/

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

Verhulst, S. G., & Sangokoya, D. (2015). *Data collaboratives: Exchanging data to improve people's lives*. Retrieved April 14, 2022, from https://sverhulst.medium.com/data-collaboratives-exchanging-data-to-improve-people-s-lives-d0fcfc1bdd9a

Verhulst, S. G., Young, A., Zahuranec, A. J., Aaronson, S. A., Calderon, A., & Gee, M. (2020). The emergence of a third wave of open data. *Open Data Policy Lab*. Retrieved April 14, 2022, from https://apo.org.au/node/311570

Williams, S. (2020). *Data action: Using data for public good*. MIT Press.

Yao, X., McCosker, A., Albury, K., Maddox, A., & Farmer, J. (2021). Building data capacity in the not-for-profit sector: Interim report. *Swinburne University of Technology*. Retrieved April 14, 2022, from https://apo.org.au/node/314477