ADM +S

BUILDING DATA CAPACITY IN THE NOT-FOR-PROFIT SECTOR

Interim Report

SWINBURNE UNIVERSITY OF TECHNOLOGY

INTERIM REPORT

XIAOFANG YAO
ANTHONY MCCOSKER
KATH ALBURY
ALEXIA MADDOX
JANE FARMER

SEPTEMBER 2021

SWINBURNE UNIVERSITY OF TECHNOLOGY

DOI: https://doi.org/10.26185/ey72-s303

Recommended citation: Yao, X., McCosker, A., Albury, K., Maddox, A., Farmer, J. (2021). *Building Data Capacity in the Not-For-Profit Sector: Interim Report*. Swinburne University: Melbourne.

Acknowledgement of Country

We respectfully acknowledge the Wurundjeri People of the Kulin Nation, who are the Traditional Owners of the land on which Swinburne's Australian campuses are located in Melbourne's east and outer-east, and pay our respect to their Elders past, present and emerging. We respectfully acknowledge Swinburne's Aboriginal and Torres Strait Islander staff, students, alumni, partners and visitors. We also acknowledge and respect the Traditional Owners of lands across Australia, their Elders, Ancestors, cultures, and heritage, and recognise the continuing sovereignties of all Aboriginal and Torres Strait Islander Nations.

The ARC Centre of Excellence for Automated Decision-Making and Society

This project was funded by the Australian Research Council Centre of Excellence for Automated Decision-Making and Society – ADM+S (CE200100005). The ADM+S aims to reduce social and economic risks in the development and implementation of automated decision-making, and to improve outcomes in transport, health care, social services and news media. The ADM+S Centre's research, training and public programs will contribute to reducing inequalities for disadvantaged groups; improving the efficiency of services and the effectiveness of public and private investment; and expanding workforce capabilities in industry, civil society, and research. The Centre will work to improve public participation in the development of automated systems. Its research will inform policy debate, and foster public confidence in responsible, ethical and inclusive automated decision-making. The Swinburne Node of ADM+S Centre's overarching aim is to reconfigure data practices, develop data and AI literacy in the community, and promote collaborative systems for data sharing and analysis that are essential for socially responsible AI and ADM.

Ethics Statement

This project was approved by Swinburne's Human Research Ethics Committee in line with the National Statement on Ethical Conduct in Human Research (Ref: 20215602-6271)

Table of Contents

ADM +S

SUMMARY

About this project

The Swinburne node of the ARC Centre of Excellence for Automated Decision-Making and Society (ADM+S) (2020-2027) is leading this project to understand and address the state of technology and data capacity in Australia's not-for-profit (NFP) sector.

Aims: The project aims to understand and build data capacity and capability *within*, and *with*, the NFP sector. This involves qualitative research using co-design processes to develop a Data Capacity Framework.

Approach: We adopt a participatory design approach and draw on local knowledge and perspectives from a range of NFP organisations to map data capacity across the sector and co-design a data capacity framework. This involves: preliminary interviews with sector representatives; co-design workshops focusing on elements of a data capacity framework; overview of exemplary use-cases.

Findings so far

The preliminary findings presented in this Interim report explore the features of data transformation among representatives of the NFP sector. The report draws on interviews with 20 people from 9 Australian NFP organisations and 13 workshop participants at the Infoxchange Group's Connecting Up conference in May 2021.

Our analysis identified five key themes: concerns with consistency in data systems and processes; development of data expertise; a preference for off-the-shelf tools but tensions in their suitability for purpose; data management strategies beyond legal compliance; and the need to expand data-driven decision making and measurement for impact.

1 Consistency in data systems and processes

- NFPs generate rich quantitative and qualitative data.
- Some of the data are not well documented and can fail to generate specific insights. Some data come from disparate sources or are lost in historical accounts or superseded systems.
- Digital transformation is ongoing, in some cases involving transferring handwritten, paper-based information into digitalised data in cloud-based systems.
- Data sharing opportunities are crucial for benchmarking to understand where NFP organisations are placed in the sector, but are difficult to establish.

2 Development of data expertise

- All participating organisations are positive about opportunities for workforce upskilling to enable better communication and analysis of data.
- Many NFPs are operating "at capacity" and any investment in data analytics must give way to maintaining the provision of essential services.
- Capability development can be empowering as it allows staff to "be in the driver's seat" and to access the necessary data more easily.
- Very few organisations have dedicated data specialists, and personnel overseeing data are not necessarily equipped with the knowledge and skills for detecting patterns and creating data-driven narratives, especially in relation to unstructured, text-based data.

3 Preference for off-the-shelf tools

- NFPs tend to opt for off-the-shelf tools that are cost-effective, presenting a shorter learning curve, and limited scope for customisation.
- Off-the-shelf tools aid data collection, data visualisation, and data analytics in the first instance, but can lack flexibility and interoperability with other systems.
- NFPs have limited time and budget for trial-and-error testing customised or bespoke technological tools.
- Collaborative enterprise structures can offer shared access to technologies and human resources and present an alternative to complex, resource-consuming data infrastructure, especially in small for-purpose organisations.

4 Data strategies beyond legal compliance

- Most organisations have sound understanding of the legal requirements regarding privacy, confidentiality, security and de-identification in the treatment of personal data.
- NFP organisations need to manage risks associated with data breaches but there are many gaps in the policies and procedures around data security.
- Comprehensive data strategies; organisation-wide cultures for data-driven insights; and ethical principles of data-driven decision making are yet to be established and extended beyond compliance with privacy laws and funder requirements.

5 Data-driven decision making and impact measurement

- NFP organisations share a consensus on the importance of data-driven impact reporting and measurement to attract funding, and inform advocacy for policy change.
- The outcomes of programs are difficult to measure, and organisations find it challenging to articulate the impact of their services via data analyses.
- Most are positive about how their services could be improved by automation but are not ready resource-wise to use AI-driven or automated decision systems.
- There is a strong desire to work towards an automated future where NFPs have the capacity to use systematic data-driven analysis, have strong and flexible data systems, and automated reporting and data visualisation process.

Next steps

Based on findings from our Phase One research activities, and a review of relevant research and grey literature, we have identified the five key elements of NFP data capacity. These include: data access, capability, infrastructure, governance and impact.

In Phase Two, we will work with NFP organisations across different domains to achieve a shared understanding of a practical data capacity framework. Through co-design workshops, we will refine the core elements of data capacity, and convert these into a set of practical steps and considerations that can help assess capacity, or measure the likelihood of success in each key dimension. This grounded understanding will inform subsequent co-development of a Data Capacity Framework that can guide data practices across the NFP sector.

ADM
+S

ABOUT THIS PROJECT

There is a growing sense of excitement and urgency around the potential for innovation that data capability can play across the not-for-profit (NFP) sector to improve social outcomes. But the challenge of doing more with data can be substantial for small and medium NFP organisations.

Aims: This project has the core aim of working collaboratively with NFP sector organisations to investigate data capacity and build capability in the sector. This will involve extensive qualitative research using participatory processes, and co-designing a Data Capacity Framework. We believe the NFP sector is willing and able to work together to share data, skills and resources to achieve social good outcomes. This collaborative approach will establish a knowledge base and help to understand the challenges and needs of the sector as it moves toward data-driven systems and automated decision making. The project is ongoing (2021-2022) and advocates for sector-wide mapping of data practices in Australia and globally with a focus on equity and inclusion.

Outcomes: Outcomes include: (a) a cross-sector data capacity and collaboration framework informed by best practices and shared principles in data use and capability-building; and (b) first steps toward a community of practice for sharing data capability knowledge, tools, experiences and best practice across the sector.

Partnerships and participation: Our principal research partner, Infoxchange, is a social enterprise with a mission to promote digitalisation for social justice, that assists organisations on their digital transformation journey. Infoxchange's 2020 survey of digital technology in the NFP sector revealed cross-sector improvements in digital infrastructure, systems and risk management, but also showed that Australian NFP organisations in general are underprepared for the advanced data analytics practices that underpin automated futures[1].

Infoxchange's **Digital Transformation Hub**,[2] funded by Lord Mayor's Charitable Foundation and Gandel Philanthropy, will help NFP organisations to build their digital capability and enable them to support service transformation and new ways of working in a post COVID-19 world. The Building Data Capacity project aims to contribute to the knowledge base needed to achieve digital and data transformation across the sector.

[1] Infoxchange Group. (2020). *Digital technology in the NFP sector*. https://www.infoxchange.org/sites/default/files/digital_technology_in_the_not-for-profit_sector_2020.pdf

[2] Infoxchange, Digital Transformation Hub, https://digitaltransformation.org.au/.

INTRODUCTION

With the increased datafication of government and corporate sectors, novel technological tools and heterogeneous data sources have been used to shape public and private services in ways that contribute to evolving data governance practices. Data driven social partnerships have been formed between government agencies and businesses to unlock the value of data by sharing datasets, exchanging data analytics expertise, and automating the process of decision making[3]. These organisations are often well-resourced in data science expertise and data management infrastructure, with strategic trajectories toward data-driven processes and decision-making.

Not-for-profits (NFPs), social enterprises, and other civil society organisations, however, are far less equipped with the capacity to systematically collect, process and analyse data generated as part of their everyday operations – with some notable exceptions. Although there is a growing sense of excitement and urgency around the need for digital transformation, small- and medium-sized NFP organisations face challenges in realising the benefits of datafication. This is often attributed to a lack resources and expertise needed to perform effective data analytics and visualisation and form data partnerships[4]. There is a reliance on commercial tools and data management platforms, and outsourcing of expertise.

However, the disparity between private, public and community sector data capability is far from straight forward.

As the first step toward building data capacity, we are investigating the state of play across the sector, reviewing best practice and international research, and working with organisations to co-design a data capacity framework.

This interim report presents preliminary findings from the scoping discussions and survey interviews with NFP and social enterprises selected from our research partner's network. The needs and challenges shared by these organisations are analysed in relation to key dimensions of the data capacity framework, including data access, data capabilities, data governance and ethics, data infrastructure, and data-driven outcomes. The report concludes with a reflection on the methods used for building data capacity for the NFP sector. It also points out future directions for the project which involves walking through and validating the framework with NFP organisations of varying sizes and domains.

[3] Susha, I., Grönlund, Å., & Van Tulder, R. (2019). Data driven social partnerships: Exploring an emergent trend in search of research challenges and questions. *Government Information Quarterly*, 36(1), 112–128. https://doi.org/10.1016/j.giq.2018.11.002

[4] Albury K., Aryani A., Farmer J., Kelly, J., McCosker A., Silva, S.; Tucker, J., Woo, J. (2021). Data for Good Collaboration: Research report. Melbourne: Swinburne Social Innovation Research Institute, Swinburne University of Technology. https://doi.org/10.26185/x93d-4v29

ADM + S

HOW ARE WE COLLABORATING WITH NFP ORGANISATIONS?

As more sectors are transformed through computation, automation and data-driven decision making, the ability to understand and use data becomes essential. Such ability is often understood in terms of data literacy. At the first level, this encompasses practical technological expertise required to find, select, understand, interpret, evaluate, and manipulate data. It also involves the critical understanding of the social contexts that relate to the use of big data, the ethical implications of data-driven decision making for vulnerable groups, and the metadata about data collection, processing, and management methodologies[5,6].

Programs and resources aiming to promote and develop *data literacy* usually focus on improving the skills and abilities of non-specialist individuals (as opposed to software developers, computer engineers and data scientists).[7] However, NFPs need more systematic and sustainable help and guidance on the best practices for training a data literate workforce within constrained budgets and human resources. Consequently, our approach focuses on **data capacity** as a more holistic term referring to all aspects that underpin or enable data-driven social outcomes.

The NFP organisations we are working with span a wide range of community, health, humanitarian and charity services. Enthusiasm for this collaboration has been consistent across small NFPs seeking efficiency gains, medium enterprises looking to consolidate and improve data systems and skillsets, and large NFPs looking toward cutting-edge solutions, automation and AI solutions. We continue to welcome new collaborators.

Participatory research: To understand data capacity across the sector and co-design a *Data Capacity Framework* we adopt a participatory design approach consisting of three groups of collaborative research activities. Our approach draws on local knowledge and perspectives from a range of NFP organisations[8,9]. This involves: a) preliminary workshops and interviews with sector representatives; b) co-design workshops focusing on elements of a Data Capacity Framework; and c) case studies (see Figure 1).

[5] Frank, M., Walker, J., Attard, J., & Tygel, A. (2016). Data Literacy—What is it and how can we make it happen? *The Journal of Community Informatics*, *12*(3). https://doi.org/10.15353/joci.v12i3.3274

[6] Fotopoulou, A. (2020). Conceptualising critical data literacies for civil society organisations: Agency, care, and social responsibility. *Information, Communication & Society*, 1–18. https://doi.org/10.1080/1369118X.2020.1716041

[7] Pedersen, A. Y., & Caviglia, F. (2019). Data Literacy as a Compound Competence. In T. Antipova & A. Rocha (Eds.), *Digital Science* (Vol. 850, pp. 166–173). Springer International Publishing. https://doi.org/10.1007/978-3-030-02351-5_21

[8] Cornwall, A., & Jewkes, R. (1995). What is participatory research? *Social science & medicine*, *41*(12), 1667-1676.

[9] Lupton, D., and Watson, A. (2021). Towards more-than-human digital data studies: Developing research-creation methods. *Qualitative Research*, *21*(4), 463–480.

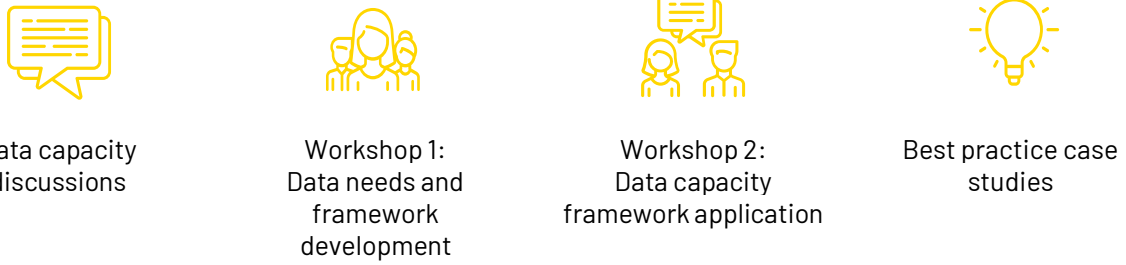

Figure 1. Project research activities

Phase 1 Phase 1 – Data capacity discussions – involves scoping discussions and interviews with representatives from across the NFP sector. These centre on organisational context and availability of data infrastructure, issues and challenges with data use and expertise, as well as goals and visions for applying data analytics and automated decision-making systems.

Phase 2 The overall aim of Phase 2 – workshops - is to contextualise the range of dimensions of data capacity. The first stage brings together participants with various professional roles and levels of data expertise. Drawing on initial scoping discussions, we will collectively define, characterise and further clarify the elements of a Data Capacity Framework as it applies to different contexts and settings. This phase will also elaborate steps for data capability building strategies relevant to participating organisations. The workshops will also test the applicability and usefulness of the Data Capacity Framework; and test the Framework as a practical strengths-based and outcomes focused tool.

Phase 3 In Phase 3 – Best practice case studies – leading-edge organisations will be selected, and in-depth interviews undertaken with key personnel to understand best practice data analytics and automation. The aim of the case study process is a) to build understanding of the applicability of the Data Capacity Framework in specific organisations and b) to assess its gaps and strengths.

These research activities are designed to build depth in our understanding of the current data practices in NFP sector in the context of digital and data transformation. Interviews and workshops will facilitate benchmarking and lay the foundations for future collaboration with NFP sector data and technology professionals. These activities will build shared understandings of self-driven and collaborative capability scaling processes that can address the specific needs and contexts of diverse NFPs

WHAT'S THE STORY SO FAR? THE NFP DATA TRANSFORMATION JOURNEY

This section presents our preliminary understanding of the features of data transformation across the NFP sector. It draws on interviews and scoping discussions with 20 people from 9 NFP organisations, mostly based in or around Melbourne, Australia, and 13 workshop participants at the Infoxchange Group Connecting Up conference in May 2021.[10]

Overview: Phase 1 discussions provide detailed information about the contexts and settings for everyday NFP sector data practices and digital transformation journeys. Using thematic analysis, we were able to group the responses into five clear key themes. These involved concerns with consistency in data systems and processes, development of data expertise, a preference for off-the-shelf tools but tensions in their suitability for purpose, moving data management strategies beyond legal compliance, and the need to expand data-driven decision making and measurement for impact.

3.1 Consistency in data systems and processes

Key points:

- NFPs generate rich quantitative and qualitative data.
- Some of the data are not well documented and fail to generate specific insights. They may be lost in historical accounts or superseded systems, or come from disparate or disorganised sources.
- Digital transformation is ongoing. In some cases, it involves transferring handwritten, paper-based information into digitalised data in cloud-based systems.
- Data sharing opportunities are crucial, but difficult to establish. In the absence of benchmarking, NFP organisations may find it difficult to understand where are placed in their sector.

Representatives from the NFP organisations surveyed reported a variety of data types generated through everyday operations. These most commonly include financial transactions data, employee data, and service delivery data. Financial transactions and employee information tend to be well captured and governed by the financial accounting department and human resources department respectively. In contrast, service data - both structured and unstructured - often fall through the cracks, with poor or inconsistent documentation, or lack of oversight from dedicated staff.

For example, one interviewee believed their organisation needed a better understanding of data to inform their provision of food services during COVID

[10] https://www.connectingup.org/

restrictions. The interviewee's organisation is beginning to digitise and make better use of data, but is also struggling to maintain operations under the increased pressures of COVID-19:

> *Some of the things that we have done is around the amount of hours that people are volunteering, so the hours of volunteering is definitely something that we collect. And then also, the number of people that come to our service. And that's been a harder one for us to actually do. We would love a system that pops up and goes, "Oh, you've been here before," but we don't. So, we definitely track each pick-up, how many pick-ups have taken place, and then we like to know what they collected as well. So, whether they collected five loaves of bread or one, or do they come every day and collect one, one, one? Or do they collect six and then never come back for a week? We really would like to know the volumes that people are collecting. (Senior manager, food service)*

A participant from adult education organisation wanted more specific data insights to optimise their course offerings:

> *We keep anonymised data on the members – so the type of members, their gender, the number of members per [branch], their general age, and also yeah, their gender. Like I was saying, so that we can keep - and understanding what sort of courses they're doing. So that's one part. We provide and establish a lot of information... We'd like to understand a breakdown of the type of members, their ages, their genders, their geography, what classes they're going to. What other things they would like, their preferences... How many people are signing up online course? How many people need to know how to get training on their phone? What sort of phones have they got? You know how many people in your council are aged over 65? How many people are bedridden? How many people are wheelchairs? How many people are willing to do courses? ... (Manager, IT, adult education)*

These interview responses indicate the rich services and operations data that organisations are or would like to generate. This data could be quantifiable measures such as volunteer hours and number of people served, or qualitative statements about personal preferences or subjective responses to surveys. Overall, interviewees expressed an ambition to capture more comprehensive data that could help organisations to more fully gauge their performance and impact over time.

The amount of data that NFPs can harness was an issue - but concerns were also expressed regarding organisational approaches for recording, documenting, and storing data. Importantly, it was recognised that disparate sources of data and loss of historical data are roadblocks for improving data quality in the context of ongoing digital transformation. This was an important issue for the staff member driving the

digital transformation processes for an organisation dedicated to youth health and wellbeing services:

> *What we are doing is looking at what we currently have and trying to improve the quality of the data because we are dealing with a lot of legacy systems and these systems are something that was initially put together 10 to 12 years back. So it's not up to date with the current technologies. And because there are so many disparate sources of data and there's obviously information silo that takes place because of that – so right now, what we're doing is we are ensuring that our main client information system, the data that's captured in there is at least compliant as per our reporting requirements. (Digital transformation officer, youth health and wellbeing)*

Due to inconsistencies in data documentation methods, a major part of the digital transformation job is to transfer handwritten notes or paper-based information onto digitalised data in cloud-based systems, and to create more systematic and consistent processes for data collection. Digitisation also helps make data more accessible for frontline staff, as it would be much easier for them to locate, enter, and query data. These are whole of organisation transformations:

> *A lot of paper-based information and putting it into a system so that it becomes readily available and useful for people. So basically, it's putting the information in the hands of people that need it, rather than sitting in file structures somewhere, and no-one knows how to find that particular article, or that particular therapy report, or that particular – because it's handwritten notes sitting in scanned documents, and you can't search it. So, that's what I mean essentially by information uplift, is bringing that information out and up, so people can actually make use of it. (Manager, business operations, disability service)*

Data collaboration can be an important means for ensuring easier and wider access to proprietary data. However, none of the NFP organisations in this study reported actively sharing their own data, or using shared or open data:

> *Even still if that [data sharing] took place, you then have historical data that wasn't built in that system, and so then it's not as comparable. And knowing how to still use data that you've had in the past but can't compare. And knowing how to present that in a way that's truthful. For instance, we have some gaps of data where we might not have collected information. Like when we moved to that custom-built software for our bread, we didn't have any software in place for two months. (Senior manager, food service)*

Data sharing opportunities are seen as crucial for benchmarking to understand where organisations are placed in the sector or against others doing similar work. As noted by one participant, this would be particularly useful for disability service providers, as data collaboration would improve coordination of support and its management. However, data loss, data storage and management are pressing problems that need to be fully addressed before data sharing and collaboration can proceed without roadblocks.

3.2 Development of data expertise

Key points:

- All participating organisations are positive about opportunities for workforce upskilling to enable better communication and analysis of data.
- Many NFPs are operating "at capacity" and any investment in data analytics must give way to maintaining the provision of essential services.
- Capability development can be empowering as it allows staff to "be in the driver's seat" and to access the necessary data more easily.
- Very few organisations have dedicated data specialists, and personnel overseeing data are not necessarily equipped with the knowledge and skills for detecting patterns and creating data-driven narratives, especially in relation to unstructured, text-based data.

Very few NFP organisations have dedicated data specialists per se, although data has permeated all aspects of their operational activities. The absence of data expertise means that organisations are unable to detect patterns and build impact narratives based on analysis and visualisations of their historical data, let alone engage in advanced data wrangling and predictive forms of data analytics.

In our study, participants who regularly undertook data analytics most often defined their roles in terms of 'managing' or 'strategic planning', across diverse organisational areas including research, finance, education, technology, business intelligence, compliance and policy, and human resources. Many came to data analysis through their own initiative, in search of better ways of reporting to executive management, Board members or funding bodies. Very few participants reported formal training in statistics or computer science, and therefore had varying levels of expertise in data use and data analysis:

None of us probably have formal data analytics training but, between kind of doing the pivot tables, count-ifs and all of that, we can start to really understand productivity levels. For me, it's really, really based off Google sheets in Excel, so being able to set

up tables or formulas to understand those, setting up the key metrics. So that's kind of the base level that I'm coming from … So we're all kind of – we know our way around sheets, but we don't have any other formal experience or skill sets. (Business strategy, youth employment service)

When asked about their perceptions about the data capability of their overall workforce, participants recognise the need to become more tech-savvy in order to enter data more accurately and use visualisations to communicate data for stakeholders. One participant from the disability service provider noted the pressing need for a more data-literate workforce as the first step to guarantee access to complete and accurate data:

Probably the biggest thing and the hardest thing is going to be the skill level of those working with data or entering data. So, a lot of the people that work in the industry are not necessarily technical savvy. They're very, very good at what they do, and very, very good at looking after people, but when it comes to turning a computer on, sometimes they panic … And that's one of the biggest challenges is like here's this great visual tool that shows us something amazing about what we've done. And they'll look at it and go, "Oh, that's a cool graph," and that's about as far as they go. And you're like, "But I need to tell you a story around this graph. " So, I think that's going to be probably one of the biggest challenges is using business speak, is selling what we're trying to do to those who don't necessarily feel like it's important. (Manager, business operations, disability service)

All participating organisations were very positive regarding opportunities to upskill workforces to manage data effectively, but recognised that this could be a complex process:

When I say training staff I mean really getting people on board with a system as opposed to manual handling, and it's just a lot of that change management stuff where we're trying to share, really champion the merit of having everything on a Cloud based system as opposed to doing everything by email or over the phone and stuff … But I do have to be also realistic and mention that there's still a bit of a skills gap. So even though it's a very user-friendly system some of the end users are not used to having everything documented in a system … that knowledge management really is very key person dependent. So it's changing that mindset as well. (Board member, leadership and governance, youth health and wellbeing)

Many organisations we engaged with are still at the early stage of scrutinising their data, applying analytical processes, and gaining insights. Strikingly, many participants reported that their organisations collect vast (and under-utilised) amounts of text-

based or qualitative data. This data either 'sits' somewhere in the system or is only being partially analysed through resource intensive manual coding. Participants observed that they knew qualitative data could yield important insights to help their work, but they lack capability for processing and analyses at scale:

> *But then, there's a lot of, I guess you would call unstructured data, which is probably where the challenge lies. And that's looking at notes around behaviours and so on, where it'll be written in either the therapist's or the support workers' style of writing, which will be inconsistent of course across, because everyone writes differently. But there's a huge amount of data there that I'm sure – It holds a lot you know. And that's the thing that I think we have a lot of, but we don't really have it, if that makes sense. And that'd be amazing, unlocking that sort of information. (Manager, business operations, disability service*)

Many NFPs are already operating at capacity in terms of offering essential services for people in need, and the tools for automated qualitative analyses – including natural language processing (NLP) and other machine learning text analysis techniques - are still in their infancy. The long-term goal of transforming services with data analytics often gives way to more urgent demands and targets:

> *At the moment we've got more people as participants in our program than we have meaningful roles to offer, so the pressure on us right now is on the one hand – in a way not all of it has hit yet, because some people are in quarantine and that's meant that the reality of not having shifts hasn't particularly hit them. So we've extended our coaching service and we've extended the peer support forums and things, but there's going to come a time where we really might struggle to keep up with the transitions. There's not a lot of work going right now externally. (Program Manager, Youth employment service)*

One one hand, NFPs are faced with the challenge of not being able to fully leverage the data they have. But this can be equally seen as an opportunity for data capacity building interventions to step in, and drive the sector towards automation and data-driven decision making. Our study indicates that many organisations are just starting to examine the data they collect, and are increasingly becoming aware that they are underprepared or lack the skills needed to undertake useful data analysis.

3.3 Preference for off-the-shelf tools

Key points:

- NFPs tend to opt for off-the-shelf tools that are cost-effective, presenting a shorter learning curve, and limited scope for customisation.
- Off-the-shelf tools aid data collection, data visualisation, and data analytics in the first instance, but can lack flexibility and interoperability with other systems in the longer term.
- NFPs have limited time and budget for trial-and-error testing customised or bespoke technological tools.
- Collaborative enterprise structures can offer shared access to technologies and human resources and present an alternative to complex, resource-consuming data infrastructure, especially in small for-purpose organisations.

The lack of capability, expertise and human resources outlined above is exacerbated by limitations in data collection, storage, and analysis infrastructure. NFPs tend to opt for off-the-shelf tools that are cost-effective and require a shorter learning curve. However, these are rarely customisable and often not designed to be fit for purpose in the NFP sector. Most participating NFP organisations and social enterprises reported that they prefer ready-to-use tools over bespoke or customisable software. No participating organisation had commissioned their own software or data solutions. Some often-cited computer systems and analytics toolkits among our participants include: Microsoft Suite (especially Excel); Google Docs; surveying tools (Survey Monkey); scheduling and rostering (When I Work); point-of-sales (Counter); project management (Folio); video conferencing (Zoom); reporting database (RiskMan); visualisation (Power BI); and customer management system (Salesforce).

These off-the-shelf tools collectively constitute the data infrastructure in many organisations. Toolkits are used to assist with operational needs, or address the needs for data collection, data visualisation, and data analytics in a preliminary manner. The transition from manual, paper-based systems to online survey tools has represented a step up the ladder of digitalisation. However, storing data in spreadsheets seems to be a crude method in itself, as indicated by one participant:

> *I mean I think moving to a 'Survey Monkey' model for students and participants is a more effective way of keeping the data than what we currently do with the manual form fill in. We store all the manual surveys at work. Everything is anonymous so people's names aren't given, so we don't record any names on that. And then it's all stored in Excel and then we have a share drive that we can share it in. But it's pretty*

basic in terms of it's stored in Excel, the participant survey research forms. (General manager, mental health service)

Participants indicated that underdeveloped data systems led to limitations in data access. Data generated and stored within siloed systems is often disrupted by software updates, and cannot communicate across systems or be aggregated for analysis or insights. This problem was noted by several participating organisations in our discussions, who raised concerns regarding the lack of interoperability between databases, the inadequacy of off-the-shelf software, and the disparate sources of data respectively:

And there's also a lot of historic stuff in there so again it predates me but I believe the current system we use, we've only been using since sort of 2017. And there's actually still a sort of back end of legacy data that is in the system but you know, there's like fields that no-ones even heard of that contain information that comes from the old database. And that's where some of the strategic stuff, the longevity of data I don't know is great, so things like the type of project for anything pre-2017 doesn't exist because it's from an old database and that didn't come across. (Manager, learning & insights, rural community development)

So, we don't have a very well organised system in this place, and that's probably because there isn't something that's tailormade to what you're doing, and to access that tailormade system is generally expensive. So, we've had lots of creative uses of low-cost software that handles certain aspects, but then they don't all talk to each other, and then the data that's within them gets forgotten about, because they're not all in one place. (General manager, food service)

Because there are so many disparate sources of data and there's obviously information silo that takes place because of that – so right now, what we're doing is we are ensuring that our main client information system, the data that's captured in there is at least compliant as per our reporting requirements. (Manager, digital transformation, youth health and wellbeing)

One technological solution to these problems would be to use tailormade software and systems. However, as much as NFPs are willing to experiment customisable technological tools, their constrained time and budget leave little room for trial and error use of these toolkits. Instead, they opt for user-friendly software that allows some degree of customisation and features essential functionalities such as record entry and report export:

We began trialling at the start of this year Volaby, which is a new software from OrangeSky, they're a non-profit organisation … they've been trying to develop an impact-tracking and volunteer schedule software that charities can use … that's probably the newest thing I've seen from a non-profit trying to do some software, but it just doesn't have enough. And as much as I want to support their mission for giving data to non-profits and helping them have that impact stuff … but it's just not there yet and requires a bit too much admin of doing this, doing that. Something goes wrong, you can't undo it; lots of stuff. And the schedule's not as comprehensive, there's no auto-tracking and stuff in various places that should be. And so, I've just decided to unfortunately let them know, "Look, I love their mission, but my mission can't be supporting your mission." And it's costing us time to try that software. (General manager, food service)

In the ideal situation, NFPs would invest in data infrastructure that allows for the development of cloud-based data storage, interoperable data systems, and customisable software based on their specific needs. However, this is not the most cost-effective option for small and medium sized NFP organisations and social enterprises who rely on donations and government funding to survive. One NFP organisation in our study received pro-bono support for data analytics from a large private sector organisation. Although the project outcome was not all successful due to lack of specificity in the insights generated by pro-bono data analysis, the collaboration provides an exemplar case for utilising a partner's resources to achieve some data analytics goals without expending much on infrastructure and systems. Another way forward could be the formation of collaborative enterprise structures, allowing NFPs to share expertise, infrastructure, and - where appropriate - data, across organisations.

3.4 Data strategies beyond legal compliance

Key points:

- Most organisations have a sound understanding of the legal requirements regarding privacy, confidentiality, security and de-identification in the treatment of personal data.
- NFP organisations need to manage risks associated with data breaches but there are many gaps in the policies and procedures around data security.
- Comprehensive data strategies, organisation-wide cultures for data-driven insights, and ethical principles of data-driven decision making are yet to be established or extended beyond compliance with privacy laws and funder requirements.

In our discussions, NFP organisations generally demonstrated comprehensive understandings of the legal implications of access, privacy, confidentiality, anonymisation and de-identification around the collection and sharing of personal data:

> *Everything's password protected. We use a Google suite, so within programs only certain people within the organisation can view data. But yeah, between those people it's viewable to all four of us. (Program manager, youth employment service)*

> *You just follow the [Privacy] Act. They're horrible to read, but they're pretty straightforward. Yeah, you only capture the data that you need, you don't capture data that's irrelevant, because it's an unnecessary risk to capture the information. (Manager, business operations, disability service)*

> *We only collect the year of birth, not the actual birthdays for example. So if that data does get lost or goes out of the organisation, it's very difficult to track people. We don't give out people's names and addresses. The analytics that we derive is anonymised so we're providing the number of males, number of females, the general ages. We're not going down to detail. We don't provide a list of names or email addresses to other organisations. (Manager, IT, adult education)*

Funder requirements regarding data collection and data sharing were prioritised:

> *We've set up a governance now where any time there's a new contract or new program that needs to start, it goes through a process where the contract manager reaches out to the data management team and we take a look at what are the requirements for this program and what kinds of systems can we put in place to support this. And all the legal considerations and the privacy considerations of this contract or for this program has been assigned to – so we've got a privacy officer and a legal and compliance manager. So they take a look at all those aspects. And when it comes to data security and everything, we've got an IT team which takes a look at those kinds of requirements. (Digital transformation officer, youth health and wellbeing)*

Participants also identified a need to manage risks associated with data breaches, which might jeopardise their efforts to protect personal and sensitive data:

> *The biggest risks, there are employees exporting data and what are the key things of the data that they have access to? A lot of them – so kind of the management team and the operational leads – they have access to kind of price sensitivity stuff which is obviously important … One thing that we do is all the team members use company*

phones so they're not logging into apps from their own phone, which is actually because we don't want to put the pressure on [trainees] to buy phone plans that have a lot of data. That's the main reason but, from the security perspective, that's actually a quite handy thing. (Business strategist, employment service)

When asked about organisational-level or internal policies or strategies around data collection and data manipulation, participants indicate that there has been little discussion around the ethical implications of using big data for decision making. That said, one of the participating NFPs noted that a data strategy had been integrated into their work but not formally recognised as such:

We don't have any kind of data strategy, which is something that I have seen thrown about a bit but I have looked at but it doesn't seem like it's got a particularly standard set of steps or processes that make up a standard data strategy. So I haven't necessarily worked out whether that's something that we should or shouldn't do or whether part way what I'm including in the data health check is really a data strategy, I mean arguably, it's going to come up with a series of recommendations and flow-on work and that's kind of a strategy. (Manager, learning & insights, rural community development)

For another organisation, the data security and governance procedures established for vendor selection acted as a quality assurance process:

I'm aware that there's data security and governance from a purely IT point of view. So, in terms of that, when we're doing vendor selection and working with the vendor and making sure they've got all the certifications and QA and all that sort of stuff, to make sure that the data that they're holding on our behalf is secure. (People and culture officer, employment service)

Participants described a range of disparate policies, practices, and procedures which operated across different departments, in order to ensure data quality and security. Few reported consistent, organisation-wide data strategies to deal with ethical issues such as transparency, accountability, and equity. There was a tendency to use privacy laws as both the first and last resort to inform data practices in the organisation.

3.5 Data-driven decision making and impact measurement

Key points:

- NFP organisations share a consensus on the importance of data-driven impact reporting and measurement to attract funding, and inform advocacy for policy change.
- The outcomes of programs are difficult to measure, and organisations find it challenging to articulate the impact of their services via data analyses.
- Most are positive about how their services could be improved by automation but are not ready resource-wise to use AI-driven or automated decision systems.
- There is a strong desire to work towards an automated future where NFPs have the capacity to use systematic data-driven analysis, have strong and flexible data systems, and automated reporting and data visualisation process.

Because of the many factors impeding data collection and analysis, NFPs and social enterprises found it particularly challenging to effectively measure and articulate the impact of their services. Several informants offered comprehensive accounts of how crucial it is for them to be able to demonstrate the need for, and impact of their services to community, industry and government stakeholders:

> *So we use it [data] as justification to develop for the stakeholder for example, to give them an insight into what's happening in their particular environment and how we can continue building that relationship to build further bespoke interventions for them. Also to collect more general data that we put into the larger pool, to be able to go and speak to industry because we are an industry voice as well so being able to speak collectively about what kind of data we're collecting. And then also to go to the higher end where we incorporate this with our original research to be able to build cases for government funding and better opportunities for ... industry to be supported through government and it's very difficult. (Manager, research and intervention, mental health service)*

> *Three different avenues that you want the data for, there's the community aspect, like a lot of what we are doing is trying to build community capacity so there is actually an element of trying to get research out that helps community groups and supports community groups across Australia in their own sort of work. There is an element of advocacy... it might be a bit strong word but ... donor facing work and government facing work, so a bit sort of broader policy-based advocacy but also trying to identify and show funders what the needs and issues are in remote Australia. (Manager, learning and insights, rural community development)*

In general, impact reporting represents a key agenda of all NFPs in our study, but a lack of quantifiable outcomes presented a significant challenge for some of the smaller-sized organisations:

> *They [charities and NFPs] often found it difficult to objectively measure outcomes. So they were able to talk about in qualitative terms about how people felt at the end of an intervention, but data – that data was often in scarce supply. So how do you make it a bit more of a measurable metric? (Business manager, employment service)*

> *There this notion of requital reporting, so if a benefactor or a philanthropic organisation gives us a sum, say, $50,000 to work on a particular project, it's incumbent on us to go back to that – to the organisation with an outcome with a report at the end of it. And I believe that … we're probably not yet good at saying well we got these funds, we did this intervention, here is the results … (Business manager, employment service)*

Impact reporting requires narratives demonstrating the effectiveness of services. Consequently, participants expressed a desire to include statistics and data visualisations as an objective and persuasive articulation of campaign, program and service outcomes:

> *What we have at the moment is okay, but I can only imagine, as we grow and develop, that we're going to need to rely on data. And we're going to have to rely on statistics because that's what the peak bodies want to see, they want to see stats and they want to see outcomes and they want to have proof points I guess, to either participate or fund. So I think that's something that I'd like to just be more open-minded to, to see what's possible. (Manager, research and intervention, mental health service)*

Beyond reporting past impact, data can be deployed to generate future-oriented insights that inform organisational strategies. However, the participants in our study suggested that NFPs often see themselves at the starting point of digital transformation, in contrast to private sector where such practices are more established. Many participants were positive regarding the potential for their services could be improved by data-driven insights (including elements of automated decision-making) but suggested their organisation was 'not ready'. As a participant noted, "*in terms of using that information and help our services to make informed decisions, we are yet to reach there*."

When invited to imagine automated futures for their organisation, participants indicated a range of interests, including: 1) discovery of patterns and insights in critical incident reports using systematic computer-driven analysis; 2) accessing data

seamlessly across systems with pre-filled data where possible, and 3) building workflows to minimise human efforts in producing reports and visualisations;

> *There must be some hidden, I believe, undiscovered insights that might be able to, if we could learn from them, we might be able to prevent some of those incidents from occurring. Especially there might be patterns, underlying patterns, we're just sort of not – people might have intuitions but it would be nice to be able to have some data to confirm that. (Quality assurance, disability service)*

> *I find automation in itself is really good for basically repeating activities. And there are a lot of activities that I see, particularly in the reporting and visualisation space, where you're doing the same thing every time, to create a graph. (Manager, business operations, disability service)*

> *Impact tracking, where someone donates and they can see where their dollar ends up - I think some donors and the public find that stuff cool. So, when they buy that bottle of Thankyou Water or whatever and they can jump on and track which village that ended up getting a tap, that sort of thing, that's particularly interesting. (General manager, food service)*

NEXT STEPS: TOWARDS A CO-DESIGNED DATA CAPACITY FRAMEWORK

Our research demonstrates that there is an urgent need for intervention to lift the NFP sector out of the current cycle of data capacity insufficiency. NFP organisations are keen to improve the use of the data they generate, to share knowledge, achieve benchmarking and potentially collaborate with others. They have a shared willingness to be part of a community of practice that is governed by sector-wide norms and best practices. The Data Capacity Framework we are working towards represents the first step in developing sector-specific guidance. These practice guidelines will help NFPs identify target areas of development, establish approaches for measuring impact outcomes, and manage the governance and ethical issues associated with data transformation and automation.

Based on findings from the Phase 1 research activities, and in line with a scoping review of the international literature (Appendix A), we have identified five key elements of a framework to NFPs' development of data capacity. Table 1 shows the working definitions of the five elements, the key questions for deliberation by NFP organisations, and the illustrative scenarios from interviews that serve to flesh out these concepts.

Five core elements identified in this first phase of the research present a whole-of-organisation picture of Data Capacity, involving:

1. **Data access**
2. **Data capability**
3. **Data infrastructure**
4. **Data governance**
5. **Data outcomes and impact**

Table 1. Key elements of data capacity

KEY ELEMENTS	WORKING DEFINITION	PROMPT QUESTIONS	ILLUSTRATIVE QUOTES
Access	The degree of ease in capturing quality data generated from everyday transactions; sourcing publicly available data, or sharing data with collaborative partner organisations	• What kinds of data do NFP organisations generate? • How is such data captured and in what format? • What data would NFPs like to have? • Do NFPs access and use external data or open data? • Do NFPs share data with other organisations?	"We mostly generate data that revolves around a young person's substance use. It also deals with their medical history... And the counselling sessions that we provide and any kind of service that we provide, that gets recorded as well on our client information system. And as I mentioned, the medical history and the substance use history and some basic outcomes and so on that we set with the young person." (Digital transformation officer, youth health)

KEY ELEMENTS	WORKING DEFINITION	PROMPT QUESTIONS	ILLUSTRATIVE QUOTES
Capability	The ability to utilise big data technologies, management systems, and open data platforms to obtain, integrate and analyse data, and carry out data prediction and application	• Who works with data in the organisation? • What are their levels of data expertise? • What are the necessary data skills for the NFP organisation in terms of technical/operational skills versus strategic and forward-thinking skills? • What is the organisational culture towards data? • Is there willingness to build data skills within the organisation? • What are the facilitators and constrains for data capability building in the NFP sector?	"Probably the biggest thing and the hardest thing is going to be the skill level of those working with data or entering data. So, a lot of the people that work in the industry are not necessarily technical savvy. They're very, very good at what they do, and very, very good at looking after people, but when it comes to turning a computer on, sometimes they panic." (Business operations, disability service)

KEY ELEMENTS	WORKING DEFINITION	PROMPT QUESTIONS	ILLUSTRATIVE QUOTES
Infrastructure	Technical means, operational mechanisms, compatible software, and analysis tools for the storage, distribution, sharing, connection, handling, and consumption of data	• What infrastructure is available in the organisation for the storage, processing, and analysis of data? • What are the services and tools used for data collection, data storage, data visualisation? • Does the organisation prefer customised or 'off the shelf' tools, systems, and software? • What are the challenges associated with these tools? • Does the organisation share data infrastructure with other organisations? What are the procedures and mechanisms for data transfer and access control? • What are the risk management techniques and security measures available?	"So, we don't have a very well organised system in this place, and that's probably because there isn't something that's tailormade to what you're doing, and to access that tailormade system is generally expensive. So, we've had lots of creative uses of low-cost software that handles certain aspects, but then they don't all talk to each other, and then the data that's within them gets forgotten about, because they're not all in one place." (General manager, food service)

KEY ELEMENTS	WORKING DEFINITION	PROMPT QUESTIONS	ILLUSTRATIVE QUOTES
Governance	Laws and regulations, industry-specific guidelines, ethical principles, data sharing contracts, internal data protection policies and practice frameworks governing the collection, use and disclosure of data	• What are the laws and regulations governing data use in the NFP sector? • How does the organisation address legal issues arising from data collection and data use? • How does the organisation deal with personal and sensitive data? • Does the organisation have data sharing agreements with other partners? • What are the internal policies for data governance in the organisation? • Who oversee the legal and ethical compliance of data use? • What are the responsibilities does data specialists have?	"We've got a number of policies regarding how data is used, what sort of data is collected, what sort of data is stored, what sort of data is shared. So we have to abide by the Vic Government, state government's directives regarding privacy and security, so we don't share data that's not allowed... We only collect the year of birth, not the actual birthdays for example... We don't give out people's names and addresses. The analytics that we derive is anonymised so we're providing the number of males, number of females, the general ages." (IT, education service)

KEY ELEMENTS	WORKING DEFINITION	PROMPT QUESTIONS	ILLUSTRATIVE QUOTES
Impact	The implications for subjects from whom data was collected, the effectiveness of data-related outputs and systems, and the service outcomes of data-driven decision making that can be measured by qualitative and quantitative metrics	• How do NFP organisations achieve responsible, ethical, trustworthy data practices and outputs? • How are data-driven outcomes shared and communicated to the public? • What are the measures for assessing the effectiveness of data capacity activities? • How can we measure the impact of social services provided by NFPs and social enterprises?	"We're going to need some sort of more compelling comprehensive case... A better case that is based on real data, not just kind of anecdotal feel-good stuff. So if we're going to go to government and say, 'Please, please minister can you give us a bucket of money'. We're all pretty upstanding individuals and we'll do the right thing. But yeah, I think there's kind of a nervousness from the board about putting the hand out without having a bit more of a, yeah a quantitative compelling case to tell." (Board member, mental health)

In the next steps, we will work with NFP organisations across different domains and help them navigate through the key considerations in building data capacity. To achieve a shared understanding, we will collaborate with NFP organisations to adjust the definitions of data capacity, convert the "prompt questions" into a set of core considerations that can help assess capacity or measure the likelihood of success in each key dimension, and move from informative evidence to evidence of application and strategy implications in relation to situated data practices under specific organisational contexts.

We recognise that participating NFP organisations value opportunities for networking with other NFP organisations working in similar or same domains. As such, our project works towards the goal of bringing together industry players, facilitating the sharing of data, and enabling collaborative benchmarking. Our workshop series will help NFP organisations identify internal opportunities for data capacity development, establish feasible goals for applying data-driven automation in services, and explore collaborative capacity building strategies. This grounded understanding will eventually inform subsequent co-development of a data capacity framework that can guide data practices across the NFP sector.

ADM
+S

ABOUT THE TEAM

Anthony McCosker is Swinburne Lead and Chief Investigator for the ARC Centre of Excellence in Automated Decision Making and Society (ADM+S) (2020-2027). His research addresses digital inclusion and participation, and the impacts and uses of new communication technologies, particularly in relation to health and wellbeing. As co-leader of the ADM+S Data Program, Anthony is leading work that seeks to establish responsible and effective data practices for an automated future.

Xiaofang Yao is a Postdoctoral Research Fellow for the ARC Centre of Excellence in Automated Decision Making and Society. Her doctoral research examines the linguistic landscapes of the Chinese diaspora in Australia. She showed how power, location and culture inform interpretation of semiotic practices in public spaces both physical and digital. Her research interests are mainly in the fields of linguistic landscapes, social semiotics, and media discourse. She has experience with ethnographic research methods including interviews, surveys and cultural observations.

Kath Albury is the co-Leader of the Digital Inclusion Program in Swinburne's Social Innovation Research Institute (SIRI), and is an Associate Investigator in the Swinburne Node of the Australian Research Council Centre of Excellence for Automated Decision-Making and Society. She previously led SIRI's Data Collaboration for Social Good research project, which was funded by the Lord Mayor's Charitable Foundation in collaboration with three Victorian NFP organisations. The present project builds on the methodologies, outcomes and outputs of the Data Collaboration for Social Good project.

Alexia Maddox is a sociologist of technology with research interests in the social impacts of technology, such as social media and digital networked technologies. She studies digital frontiers and generally conducts community studies with stigmatised populations using technology to create and connect in emerging spaces online. Her book, 'Research Methods and Global Online Communities: a case study' presents an approach to mixed-methods research and is written to support postgraduate and early career researchers in exploring these evolving social spaces through a myriad of techniques.

Jane Farmer is Director of the Social Innovation Research Institute at Swinburne. She has a distinguished track record of research in rural health services, innovations in health workforce and technology, community participation, coproduction and social enterprise. Addressing disadvantage and social and health inequity is central to her research activity.

Kimberlee Weatherall is a Professor of Law at the University of Sydney, and a Chief Investigator with the ARC Centre of Excellence for Automated Decision-Making and Society. She specialises in issues at the intersection of law and technology, as well as intellectual property law. Her current research focuses on the law relating to the collection, ownership, use and governance of data about and related to people, including privacy law, with the goal of ensuring that data collection, use and linkage, and data and predictive analytics are developed in ways that are beneficial to people and society.

Julia Stoyanovich is Associate Professor of Data Science at the Center for Data Science, and also Director of the Center for Responsible AI (R/AI), both at New York University. She is a recipient of an NSF CAREER award and of an NSF/CRA CI Fellowship. Julia's research focuses on responsible data management and analysis practices: on operationalizing fairness, diversity, transparency, and data protection in all stages of the data acquisition and processing lifecycle.

Jake Goldenfein is a Senior Lecturer at Melbourne Law School. He has been a researcher at Cornell Tech, Cornell University and New York Law School, and the Swinburne Institute for Social Research in the fields of media and communications history and theory, intellectual property, communications policy, privacy and media law. He is an admitted lawyer in Australia, and previously practiced as a solicitor in an international firm in the areas of privacy and administrative law.

References

Carmi, E., Yates, S. J., Lockley, E., & Pawluczuk, A. (2020). Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation. *Internet Policy Review*, *9*(2). https://doi.org/10.14763/2020.2.1481

Connell, E., Bell, S., MartinRogers, N., & Souvenir, N. (2020). Data Parties: Giving the Community Tools to Use East Metro Pulse Survey Data. In F. Ridzi, C. Stevens, & M. Davern (Eds.), *Community Quality-of-Life Indicators* (pp. 185–198). Springer International Publishing. https://doi.org/10.1007/978-3-030-48182-7_10

Fusi, F., & Feeney, M. K. (2020). Data sharing in small and medium US cities: The role of community characteristics. *Public Administration*, *98*(4), 922–940. https://doi.org/10.1111/padm.12666

Murray, B. (2015). *Data Walks: An Innovative Way to Share Data with Communities* (p. 42). Urban Institute.

Namisango, F., Kang, K., & Beydoun, G. (2021). How the Structures Provided by Social Media Enable Collaborative Outcomes: A Study of Service Co-creation in Nonprofits. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-020-10090-9

Ontario Nonprofit Network. (2015). *Towards a data strategy for the Ontario nonprofit sector*. Ontario Nonprofit Network. https://theonn.ca/wp-content/uploads/2015/07/Towards-a-Data-Strategy-for-Ontario-Nonprofit-Sector_ONN_Final_2015-07-13.pdf

Portwood, S. G., Shears, J. K., Eichelberger, C. N., & Abrams, L. P. (2010). An Institute for Social Capital: Enhancing Community Capacity Through Datasharing. *Child Indicators Research*, *3*(2), 261–273. https://doi.org/10.1007/s12187-009-9052-8

Qlik.com. (2020). *Developing a data literate workforce: A strategy and framework for the enterprise*. https://www.qlik.com/us/-/media/files/resource-library/global-us/register/whitepapers/wp-developing-a-data-literate-workforce-en.pdf

Tomašev, N., Cornebise, J., Hutter, F., Mohamed, S., Picciariello, A., Connelly, B., Belgrave, D. C. M., Ezer, D., Haert, F. C. van der, Mugisha, F., Abila, G., Arai, H., Almiraat, H., Proskurnia, J., Snyder, K., Otake-Matsuura, M., Othman, M., Glasmachers, T., Wever, W. de, … Clopath, C. (2020). AI for social good: Unlocking the opportunity for positive impact. *Nature Communications*, *11*(1), 2468. https://doi.org/10.1038/s41467-020-15871-z

Appendix: Benchmarking the dimensions of data capacity

Based on our preliminary conversations with representatives from NFPs, we suggest that the features of the data transformation journey across the sector can be mapped against five primary data capacity dimensions: (1) data access, (2) data capability, (3) data infrastructure, (4) data governance, and (5) data outcomes. When developing and evaluating data practices relevant to their contexts and specialised domains, NFPs are encouraged to think through these five components, and how they interact and are supported across the whole organisation.

These data capacity dimensions are situated in the overarching principles of responsible, ethical, and transparent data use, including the use of data within artificial intelligence and automated decision-making systems[11]. In this section, we contextualise our research findings via a preliminary review of global research on the dimensions of data capacity. By engaging with relevant scholarly research and grey literature, we deepen our understanding of current data practices across the sector, and investigate how this can be situated within broader concerns associated with enabling data-driven operations and embedding automated decision-making systems.

A.1 Access

NFPs generate rich data as part of their service delivery, but the quality and consistency of such data is often compromised. Customer-facing staff and volunteers are charged with the responsibility of collecting first-hand service data, but their limited data capability can result in data loss or data documented through paper-based medium, which then presents significant challenges for data digitalisation. These issues are essentially attributed to lack of access to quality and reliable data, which represents the first point of consideration for building data capacity.

In the public sector, access is often understood in terms of the degree of ease in sourcing, obtaining and sharing quality data subject to due authorisation processes[12]. Accordingly, international open data movements and Australia-wide legislative reforms have aimed at ensuring easy and safe access to shared public data[13]. However, data accessibility remains a substantial challenge for NFP and community-based

[11] 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*. https://doi.org/10.1007/s10676-020-09577-5

[12] Australian Bureau of Statistics. (2021). Managing the risk of disclosure: The five safes framework https://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/1160.0Main%20Features4Aug%202017

[13] Australian Government. (2018). New Australian Government Data Sharing and Release Legislation: Issues paper for consultation. https://pmc.gov.au/resource-centre/public-data/issues-paper-data-sharing-release-legislation

organisations who are struggling with capturing internal data generated from everyday transactions, let alone making the best use of publicly accessible data or engaging in collaborative data sharing with partner organisations.

There has been some criticism of the extent to which open public datasets are useful or open at all[14]. Yoon et al. (2018), in their research on community data use, report that instead of general national-level demographic data civil society organisations need access to local neighbourhood data which proves far more difficult to collect[15]. A possible solution to the problem is to engage data intermediaries, as the agents positioned between two other organisations in a data supply-demand chain that facilitate the use and reuse of data, to support NFPs in collecting data that accurately represent community profiles.

In addition, limited and uneven access remains a barrier for small community-based organisations despite the promises of open data and their public benefits. Not all data are publicly available or equally accessible due to their proprietary nature, and available data on public repositories tend to be siloed and not in a format that is readily usable or interoperable, especially for those NFPs who have compromised digital infrastructure and constrained financial and human resources to enable such access[16].

Importantly, providing equitable data access for NFPs needs to be a collective effort involving different stakeholders in the data ecosystem, including local governments, data providers, data repositories, data intermediaries, and educational institutions[17]. The issue of inadequate access to relevant and high-quality data can potentially be addressed by engaging the private sector in reciprocal relationships to enable the exchange of datasets, data expertise, and technological resources[18]. The NFP sector has been slow to adapt and take up data-driven systems, but by collaborating with public sector and other service providers they might have more opportunities to collect, share and improve access to the data they can use to support their operations and services[19].

---

[14] Wang, V., & Shepherd, D. (2020). Exploring the extent of openness of open government data–A critique of open government datasets in the UK. *Government Information Quarterly*, *37*(1), 101405.

[15] Yoon, A., Copeland, A., & McNally, P. J. (2018). Empowering communities with data: Role of data intermediaries for communities' data utilization. *Proceedings of the Association for Information Science and Technology*, *55*(1), 583–592. http://doi.org/10.1002/pra2.2018.14505501063

[16] Yoon, A., & Copeland, A. (2020). Toward community-inclusive data ecosystems: Challenges and opportunities of open data for community-based organizations. *Journal of the Association for Information Science and Technology*, *71*(12), 1439–1454. http://doi.org/10.1002/asi.24346

[17] 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).

[18] Holton, J. (2018). *The private sector to the public sector's rescue: Why private organizations enter crisis response data collaboratives* [Master's Thesis, Leiden University]. Leiden University Student Repository. https://hdl.handle.net/1887/58629

[19] Thinyane, M., Goldkind, L., & Lam, H. I. (2018). Data Collaboration and Participation for Sustainable Development Goals—A Case for Engaging Community-Based Organizations. *Journal of Human Rights and Social Work*, *3*(1), 44–51. https://doi.org/10.1007/s41134-018-0047-6

A.2 Capability

It is often found that within NFP organisations personnel overseeing data strategies are not necessarily equipped with the knowledge and skills for analysing and making best use of data insights. They might have backgrounds in statistics and computer sciences or have the ability to process text-based qualitative data and conduct advanced data analytics. These issues can be understood in terms of data capabilities, which are seen as a valuable resource whose function improves individual and organisational productivity and strategic decision-making[20].

From a business analytics or business insights perspective, NFP organisations' data capability can be seen as consisting of three elements, including value capabilities, dynamic capabilities, and competitive capabilities. Value capabilities are essential for NFPs to deliver public benefits and includes employee competences, infrastructural resources, and management structures. Dynamic capabilities allow organisations to develop new opportunities through innovation, learning and networking. Competitive capabilities enable the NFP sector to stay afloat in the digital transformation through the deployment of data strategies for targeted and improved services.

Lin and Kunnathur (2019) further suggest that big data capability is both operational and dynamic, as big data practices not only involve "operational routines of identifying, collecting, storing, and analysing" data, but also generate knowledge about "how to create, update, and reconfigure resources, assets, and competencies needed for everyday transactions (p.50)[21]. In this sense, an organisation's data capability is dependent on the combined and strategic use of various resources including a collection of datasets, skills, competencies, and cultural values. The key, therefore, lies in how to identify valuable data within large quantities of data flows and capitalise for the purpose of accomplishing social good outcomes.

In the broadest sense, data capability denotes the ability to use big data technologies, management systems, and open data platforms to obtain, prepare and integrate data, and carry out exploratory, descriptive, inferential and predictive analysis, to improve service efficacy and aid or even automate decision-making[22]. It may be measured in terms of workforce skills, system readiness, and organisational culture. Big data capabilities can be used to characterise an organisation's willingness and preparedness

[20] Zeleti, F. A., & Ojo, A. (2018). Capability Model for Open Data: An Empirical Analysis. *Proceedings of the 11th International Conference on Theory and Practice of Electronic Governance*, 403–411. https://doi.org/10.1145/3209415.3209492

[21] Lin, C., & Kunnathur, A. (2019). Strategic orientations, developmental culture, and big data capability. *Journal of Business Research*, *105*, 49–60. https://doi.org/10.1016/j.jbusres.2019.07.016

[22] Zhang, A., & Lv, N. (2021). Research on the Impact of Big Data Capabilities on Government's Smart Service Performance: Empirical Evidence from China. *IEEE Access*, 1–1. https://doi.org/10.1109/ACCESS.2021.3056486

to enhance consistency in data collection, acquire high-tech tools and practices, and deploy human resources to support strategic decisions[23].

In practice, however, data capability has been used interchangeably with data ability, data literacy, and data competence to describe a specific area of data expertise such as information processing capability, data engineering, data visualisation, and predicative analytics[24]. For this research project, we understand capability as an essential component of data capacity; whereas capacity is an all-encompassing term referring to the overall abilities and extent to which an organisation can make use of data, capability is more specifically concerned with the skills and competences of professional roles established to work with data.

A.3 Infrastructure

NFPs tend to opt for off-the-shelf tools that are cost-effective, to some extent customisable and with a shorter learning curve. However, data generated and stored within these siloed systems is often disrupted by software updates and cannot always communicate across systems or be aggregated for analysis or insights. These are infrastructural issues that require NFPs to consider the assemblage of "technological objects, standards, values, administrative procedures, and organisational work, all of which involve myriad people, institutions, technologies, policies, legalities, and financial arrangements to build, repair, maintain and reconstruct" the flows of goods, ideas, and services (Williamson, 2018, p.5)[25].

It is important to note that infrastructure is not about discrete systems but about networks and gateways comprised of shifting relations among databases, protocols, classification systems, procedures, processes, user interfaces and other elements involved in the making and use of data[26]. As such, data infrastructure represents the technical means and operational mechanisms for the storage, sharing, and consumption of data across networked platforms, and depends on compatible software, analysis tools and data commons to enable the distribution, connection, and handling of big data.

---

[23] Shaqrah, A., & Alzighaibi, A. (2021). Linking knowledge management capabilities and the mediating role of the big data capabilities for enterprise value-adding processes. *VINE Journal of Information and Knowledge Management Systems*. https://doi.org/10.1108/VJIKMS-05-2020-0087

[24] Dubey, R., Gunasekaran, A., & Childe, S. J. (2019). Big data analytics capability in supply chain agility: The moderating effect of organizational flexibility. *Management Decision*, *57*(8), 2092–2112. https://doi.org/10.1108/MD-01-2018-0119

[25] Williamson, B. (2018). The hidden architecture of higher education: Building a big data infrastructure for the 'smarter university.' *International Journal of Educational Technology in Higher Education*, *15*(1), 12. https://doi.org/10.1186/s41239-018-0094-1

[26] Gray, J., Gerlitz, C., & Bounegru, L. (2018). Data infrastructure literacy. *Big Data & Society*, *5*(2), 1–13. https://doi.org/10.1177/2053951718786316

Discussions of data infrastructure often centre around data sharing initiatives and shared data platforms. For example, the trusted data sharing framework proposed by the Singaporean Government entails a set of technical considerations including: 1) procedures around the maintenance of encryption protocols for sensitive data in databases, in memory and in system interfaces; 2) defensive mechanisms such as network intrusion detection systems, secure standardised network protocols, redundancy plans or replication backups, and entry control into physical facilities; 3) identity, entitlement and access management in relation to databases, programs, operating systems, application interfaces, servers and devices; 4) documentation of incident management and system recovery processes, data restoration and secure disposal, data input and output integrity auditing; 5) ongoing risk management by establishing risk register, evaluating changes in activities, technologies and systems, categorising data based on value and criticality, and implementing monitoring system to automatically analyse security logs[27].

Another example is shared data platforms, which have been established to provide overarching data infrastructures for querying, browsing, analysing, and processing digital contents by linking up distributed data repositories. In the education domain, there have been proposals for developing data platforms based on shared standards and interoperability, to create markets for data-driven education services and products. An illustration is the National Schools Interoperability Program (NSIP) which supports "standards-based system integration, data sharing between organisations and jurisdictions, reuse of infrastructure, and data aggregation and synchronisation" (Sellar, 2017, p.347)[28]. Castiglione et al. (2017) presents a prototype of Cultural Heritage Information System (CHIS), which is a scalable cloud-based infrastructure with the architecture of a data source layer, a data storage and management layer, and a data processing layer, capable to collect data from heterogeneous sources, managing big data resources, and providing customised data based on preferences, and conducting data analytics in the cultural heritage sector[29].

While it is generally recognised that NFPs are poorly equipped with the procedures and systems to leverage data, there is limited understanding of the status quo of data infrastructure preparedness in the Australian NFP sector. An extensive report by Infoxchange revealed that the NFP sector in Australia has not yet prepared for advanced data analytics for automated futures, despite improvements in digital

[27] Infocomm Media Development Authority of Singapore, & Personal Data Protection Commission. (2019). *Trusted Data Sharing Framework*. https://www.imda.gov.sg/-/media/Imda/Files/Programme/AI-Data-Innovation/Trusted-Data-Sharing-Framework.pdf

[28] Sellar, S. (2017). Making network markets in education: The development of data infrastructure in Australian schooling. *Globalisation, Societies and Education*, *15*(3), 341–351. https://doi.org/10.1080/14767724.2017.1330137

[29] Castiglione, A., Colace, F., Moscato, V., & Palmieri, F. (2018). CHIS: A big data infrastructure to manage digital cultural items. *Future Generation Computer Systems*, *86*, 1134–1145. https://doi.org/10.1016/j.future.2017.04.006

infrastructure and systems and risk management[30]. Indeed, with the ever-growing proliferation of cloud computing and storage services, minimal investment in data infrastructure might be possible or even beneficial for directing resources and budget to delivering much-needed services for the public, if NFPs have fully capitalised on the offerings of data sharing platforms.

A.4 Governance

It is important to note that access, capability, and infrastructure need to be governed by legal and ethical data principles and guidelines to guarantee transparent and responsible data practices in the NFP sector. Data strategies and cultures for data-driven insights at the organisational level are also essential for ensuring that NFPs employees have clarity in their dealing with data. Laws and regulations, ethical principles, and organisational data policies collectively constitute the governance structure, which represents another essential dimension of data capacity. Although there is awareness and recognition of the value of data, comprehensive internal data policies have yet to be established or extended beyond compliance with privacy laws and funder requirements.

As the report released by the Australian Charities and NFPs Commission (ACNC) suggests, Australian NFPs collect and store a wide range of personal and transaction data for purposes of providing effective services, managing clients, volunteers, and sponsors, and supporting funding applications[31]. To mitigate risks with handling sensitive data, it is advised that NFP organisations comply with federal and state level laws such as Privacy and Data Protection Act 2004, Health Records Act 2001 and Victorian Data Sharing Act 2017, and act with reasonable care and due diligence for public good in accordance with Australian Privacy Principles[32] (APPs) when setting up policies and processes for managing personal data.

National and international data-related legislation, frameworks and standards have been established that outline rules and restrictions on data sharing, but these mainly apply to public sector data. For example, the Victorian Protective Data Security Framework has detailed ongoing monitoring and assurance activities for data security through compulsory reporting processes[33]. The Trusted Data Sharing Framework from Singapore suggests that in addition to laws and regulations, industry-specific

[30] Infoxchange Group. (2020). *Digital technology in the NFP sector*. https://www.infoxchange.org/sites/default/files/digital_technology_in_the_not-for-profit_sector_2020.pdf

[31] Australian Charities and Not-for-profits Commission. (2017). *Managing people's information and data*. https://www.acnc.gov.au/file/651/download?token=-_6yGNvF

[32] Office of the Australian Information Commissioner. (2021). https://www.oaic.gov.au/privacy/australian-privacy-principles/read-the-australian-privacy-principles/

[33] Office of the Victorian Information Commissioner. (2021). https://ovic.vic.gov.au/wp-content/uploads/2020/02/Victorian-Protective-Data-Security-Framework-V2.0.pdf

guidelines, contracts between data sharing parties, as well as intellectual property rights, copyright, and databased rights, organisations should devise data protection policies and practices for internal use and compliance covering aspects such as data management, data protection and data use standards. Informed consent, accountability and openness, care and goodwill should be upheld for collection, use and disclosure of personal data.

As discussed, legal compliance requirements have primarily focused on managing risks associated with the collection, use and sharing of personal data. The broader range of issues related to reliability, safety and trustworthiness in data practices and artificial intelligence systems, on the other hand, must be dealt with by adhering to ethical principles. One often-cited example of such ethics frameworks is the Data Ethics Canvas[34] developed by Open Data Institute (ODI), which helps individuals and organisations to identity and manage ethical issues in relation to data sources, purposes of data use and implications on the publics before making any data decisions.

In addition, Data Ethics Decision Aid (DEDA) can help government agencies walk through the ethical decisions to be made in the application of algorithms and the development of data projects. The DEDA involves sets of questions to be used as guidelines and a decision support for data related considerations involving algorithms, data source, anonymisation, visualisation, sharing, reusing, and repurposing, as well as general considerations including responsibility, communication, transparency, privacy, and bias[35]. These questions have been shown to assist participants to clarify their understanding of ethical challenges in their everyday data practices within their institutions[36].

More importantly, data ethics can be used to guide the development of reliable systems, safety and trustworthy certification through sound software engineering practices, proven business management strategies, and independent oversight by governmental and non-governmental bodies[37]. The key, however, lies in the extent to which these guidelines and principles are reflective of the realities of the Australian NFP sector. Also at issue, is where NFPs are placed in terms of applying these principles to practices, given that they have limited access, capability, and infrastructure to best leverage data for automated futures. Therefore, NFP organisations need to consider the legislative and ethical data environment in which they operate when approaching the data capacity building agenda.

[34] https://theodi.org/article/the-data-ethics-canvas-2021/

[35] https://dataschool.nl/en/deda/

[36] 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*. https://doi.org/10.1007/s10676-020-09577-5

[37] Shneiderman, B. (2020). Bridging the Gap Between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-centered AI Systems. *ACM Transactions on Interactive Intelligent Systems*, *10*(4), 1–31. https://doi.org/10.1145/3419764

A.5 Impact

Ensuring data access, building data infrastructure, increasing data capability, and establishing legal and ethical environments are important for developing comprehensive data capacity. NFP organisations need to measure the social impact of services and activities and the organisational data capacity so that they can allocate constrained budgets and human resources more effectively, target at areas where data makes the biggest difference in improving service coverage and articulate more convincingly the impact of the charitable campaigns for sponsors and funders. We therefore understand impact as a final key dimension of data capacity.

In the NFP context, impact is often understood as referring to the social impact of services and activities of NFPs in terms of the difference they make in the world. There are specific metrics for measuring the outcomes of the social and humanitarian services provided by NFPs. For example, Higgins et al. (2014) suggest that crucial indicators for impactful services for include: easier access to services within the neighbourhoods (such as occupational health, alcohol management, hazards removal, and social care for older persons), improved identification of vulnerable individuals, increased reuse of demographic data, and increased delivery of collaborative services between partnering NFP organisations or government agencies[38].

Depending on the business domains of NFPs and social enterprises, other ways of quantifying the impact and assessing the effects of service offerings on individuals and the society might include measurements such as lunchbox pickup rates, levels of housing related illness, the elderly fall rate, youth early intervention rates, and number of educational programs. Importantly, however, to measure the social impact of services, NFPs must firstly have well-maintained qualitative and quantitative activity-related data and advanced data analytics and communication capability that allows comparisons between historical and current datasets and generation of impact reports.

In addition, NFPs also seek to measure the impact or outcomes of their organisational data capacity, by benchmarking it against industry-wide data practices and standards. As collectors of personal and operation data, NFPs make decisions that have consequences for the volunteers, stakeholders, and communities[39]. Therefore, their data practices and associated outputs should be responsible, ethical, trustworthy, and democratic to justify the potential risks and impacts on these subjects. In terms of

[38] Higgins, E., Taylor, M., Lisboa, P., & Arshad, F. (2014). Developing a data sharing framework: A case study. *Transforming Government: People, Process and Policy*, *8*(1), 151–164. https://doi.org/10.1108/TG-02-2013-0007

[39] D'Ignazio, C., & Bhargava, R. (2015). Approaches to Building Big Data Literacy. *Bloomberg Data for Good Exchange Conference*, 1–6.

measuring data capacity, Kennedy et al. (2016) argues that the effectiveness of data visualisation can be evaluated based on technical factors such as memorability, speed, accuracy of participant recall, and consistency of comprehension, as well as contextual and sociocultural factors including beliefs, opinions, emotions, confidence and skills[40].

Jeble et al. (2019) in particular note that one of the significant impacts of big data and predicative analytics is that it can improve the effectiveness of humanitarian missions by creating social capital for NFPs and attracting support in disaster management[41]. Specifically, they argue that metrics for measuring the performance of humanitarian efforts should be designed in line with organisational missions of NFPs, and mostly likely include the measurements of reliability, responsiveness, agility, and cost. Gupta et al. (2018) provides a comprehensive account of how big data and analytics can be applied and help NFPs achieve societal impact in healthcare, disaster response, resource management and sustainability.[42]

Other aspects of organisational capacity might be measured by the extent to which knowledge and awareness, social capital, and opportunity and impact have been increased, expanded, or heightened through data-driven decision-making systems[43]. Knowledge and awareness refer to the working mechanisms of data systems, holistic understanding of data issues, and access to community data; social capital can be quantified based on enhanced reputation and increased influence within the NFP sector; opportunity and impact can be understood as increased competitiveness for grants, expanded ability to serve clients and improved problem-solving processes.

[40] Kennedy, H., Hill, R. L., Allen, W., & Kirk, A. (2016). Engaging with (big) data visualizations: Factors that affect engagement and resulting new definitions of effectiveness. *First Monday*, *21*(11). https://doi.org/10.5210/fm.v21i11.6389

[41] Jeble, S., Kumari, S., Venkatesh, V. G., & Singh, M. (2019). Influence of big data and predictive analytics and social capital on performance of humanitarian supply chain: Developing framework and future research directions. *Benchmarking: An International Journal*, *27*(2), 606–633. https://doi.org/10.1108/BIJ-03-2019-0102

[42] Gupta, A., Deokar, A., Iyer, L., Sharda, R., & Schrader, D. (2018). Big Data & Analytics for Societal Impact: Recent Research and Trends. *Information Systems Frontiers*, *20*(2), 185–194. https://doi.org/10.1007/s10796-018-9846-7

[43] Nowell, B., & Foster-Fishman, P. (2011). Examining Multi-Sector Community Collaboratives as Vehicles for Building Organizational Capacity. *American Journal of Community Psychology*, *48*(3–4), 193–207. https://doi.org/10.1007/s10464-010-9364-3