



# Responsible data stewardship

---

March 2023

# Contents

Executive summary	2
Chapter 1: Ways of thinking about how data should be used	3
Chapter 2: Responsibility	11
Chapter 3: Responsible data practices and interventions	18
Chapter 4: What we mean by responsible data stewardship	28
Chapter 5: What's next?	31
Endnotes	32
Methodology	32
Acknowledgements	33

## About

This report has been researched and produced by the Open Data Institute, and published in March 2023. This report was written by Joe Massey, Sasha Moriniere and Ed Parkes (Emerging Field), with supporting contributions from Yusuff Adigun, Jane Crowe, and Jack Hardinges.

If you want to share feedback by email or would like to get in touch, please contact the Responsible data stewardship project lead Joe Massey at [Joe.Massey@theodi.org](mailto:Joe.Massey@theodi.org).

# Executive summary

In response to rising awareness of the harms and negative impacts of data, different concepts have emerged that represent new ways of thinking about how it *should* be used.

This report sets out the findings of research undertaken by the Open Data Institute (ODI) between June 2022 and March 2023, supported by the Patrick J. McGovern Foundation. We set out to document and explore these concepts, with a particular focus on ‘responsible data’, and to use the findings to develop a normative interpretation of responsible data stewardship that we could use in our work. The research consisted of a literature review and expert interviews, involving 18 organisations from across the public, private and third sectors, in Africa, Europe, North and South America.

In [Chapter 1](#) we discuss some of the prominent concepts we came across in our work, including ‘data stewardship’, ‘data ethics’, ‘data justice’ and ‘data for good’ and examine how these narratives have helped people to think about how data should be collected, used and shared.

We explore the concept of ‘responsible data’ in detail in [Chapter 2](#). We describe how it’s used by civil society organisations, governments, international NGOs and large technology organisations. We conclude that it’s used inconsistently and with varying depth of thought, and that its meaning differs significantly based on geographical and cultural context.

What does being responsible with data mean in practice? In [Chapter 3](#) we investigate different ways that the concept has shaped behaviours, policies and processes in the real world. We found examples of efforts designed to protect privacy, address biases and enable participation in data processing. We also document interventions that have been developed to support and compel organisations to use data responsibly – including frameworks, technologies, training and legislative action.

In [Chapter 4](#) we set out the need for a normative interpretation of responsible data stewardship to use in our work. We ultimately articulate it as ‘an iterative, systemic process of ensuring that data is collected, used and shared for public benefit, mitigating the ways that data can produce harm, and addressing how it can redress structural inequalities’.

Finally, in [Chapter 5](#) we talk about the next steps in our research agenda, including our intent to operationalise the findings from this research and work with others to help bring about more responsible data stewardship.

# Chapter 1:

## Ways of thinking about how data should be used

The right kind of access to data is vital in tackling the big challenges we face in society – from the earlier detection and treatment of disease to reducing pollution in urban spaces. Data also has an important role to play in driving economic growth by supporting the creation of new technologies, products and services.

The transformative power of data is best explained through examples like [The Human Genome Project](#), undertaken between 1990 and 2003. Led by the US Government’s National Institute of Health, the Project made data on DNA sequencing available within 24 hours of its discovery, which not only saved lives, but also generated \$796 billion in economic impact and supported over 300,000 jobs in 2010 alone.<sup>1</sup> [Safetipin](#) is a social organisation based in India working to make public spaces safer and more inclusive for women by collecting data via three mobile phone applications to generate a safety score for users to make safe and informed decisions about their mobility. Safetipin is now being used across 50 cities in India and beyond including Buenos Aires, Mombasa, and Cape Town.<sup>2</sup> The data from Safetipin has been used in Delhi to identify 7,438 ‘dark spots’ around the city, forcing the Government to fit LED lights in these areas to improve women’s safety.<sup>3</sup>

However, data and related technologies can also cause harm, including through automating decisions that need a human touch, or embedding existing biases and inequities. Scandals like [Edward Snowden’s revelations](#) or the [Cambridge Analytica scandal](#) attracted considerable public attention, and have impacted public trust in technology,<sup>4</sup> acting as catalysts for reflection on the ways that we should conceptualise and approach data. Both of these scandals exposed the extensive collection of personal data by both governments and private companies, with limited safeguards in place to ensure data was and is used in a safe and secure way. Shoshana Zuboff describes this phenomenon whereby corporations conduct widespread collection and commodification of personal data as ‘surveillance capitalism’.<sup>5</sup>

---

<sup>1</sup> Battelle Memorial Institute, Tripp, S. and Greuber, M. (2011), ‘[Economic Impact of the Human Genome Project](#)’

<sup>2</sup> Aapti Institute, Mohamed, S, (2020), ‘[Stewarding Data for Safe & Inclusive Cities](#)’

<sup>3</sup> Hindustan Times, Goswami, S. (2017), ‘[Delhi to be free of dark spots by January, claims AAP Government](#)’

<sup>4</sup> Doteveryone, Miller, C, Kitcher, H, Perera, K, Abiola, A, (2020), ‘[People, Power and Technology: The 2020 Digital Attitudes Report](#)’

<sup>5</sup> Zuboff, S. (2019), ‘The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power’

Similarly, social media giants and other technology companies have sought to bring connectivity to the [‘next billion users’](#) in the Global South, expanding the number of data subjects that feed these extractive models.<sup>6</sup> In 2016 Facebook launched Free Basics,<sup>7</sup> a free, limited internet service for developing markets. It gave Facebook control over the internet experience, locking users into using a specific set of predetermined websites. While there has been pushback in various countries, Facebook Free Basics is still active in over sixty countries. Renata Avila describes this phenomenon as ‘Digital colonialism’,<sup>8</sup> ‘the new deployment of a quasi-imperial power over a vast number of people, without their explicit consent, manifested in rules, designs, languages, cultures and belief systems by a vastly dominant power.’

AI and machine learning algorithms are increasingly being used to make decisions – including decisions about us. These algorithms, and the data that underpins them, have been shown to reinforce problematic biases. Examples like those raised by Safiya Noble, where a Google image search for ‘black girls’ showed photos of gorillas,<sup>9</sup> or the work of Joy Buolamwini to expose how facial recognition struggles to read Black faces<sup>10</sup> demonstrate the discriminatory impacts that biased datasets and algorithms can have on society. There is also emerging awareness about algorithmic amplification of certain content online. Some platforms’ algorithms and business models are designed in a way that amplifies disinformation, extremism, and hate speech narratives, which can have real-life consequences.<sup>11</sup> These algorithmic processes remain opaque for users, impacting how people understand the world.<sup>12</sup>

The negative impacts of data and related technologies on the environment are also becoming better understood through increasing interest from research, political, and advocacy groups.<sup>13</sup> There is a growing awareness that some components of the infrastructure of the Internet are not environmentally sustainable.<sup>14</sup> The energy consumption of data centres has outgrown the airline industry,<sup>15</sup> and a single data centre can consume the equivalent of 50,000 homes, with most of this energy coming from

---

<sup>6</sup> The Guardian, Solon, O, (2017), [‘It’s digital colonialism’: how Facebook’s free internet service has failed its users’](#).

<sup>7</sup> Nothias, T. (2020), [‘The Rise and Fall... and Rise Again of Facebook’s Free Basics: Civil Society and the Challenge of Resistance to Corporate Connectivity Projects’](#).

<sup>8</sup> Internet Health Report, Avila, R, (2018), [‘Resisting digital colonialism’](#)

<sup>9</sup> Noble, S. (2017), [‘Algorithms of Oppression’](#)

<sup>10</sup> Algorithmic Justice League (n.d), [‘What Is Facial Recognition Technology?’](#)

<sup>11</sup> New York Times, McCabe, D, (2021), [‘Lawmakers Target Big Tech ‘Amplification.’ What Does That Mean?’](#)

<sup>12</sup> Big Data and Society, Burrell, J, (2016), [‘How the machine ‘thinks’: Understanding opacity in machine learning algorithms’](#)

<sup>13</sup> ICT works (2020), [‘Digital Technologies Are Part of the Climate Change Problem’](#)

<sup>14</sup> CartONG (2022), [‘GeONG 2022: 2022 GeONG - RT: Becoming more sober w/ IM tools & approaches: the avenues to explore for aid actors’](#)

<sup>15</sup> The ODI, Snaith, B, (2023), [‘Data centres, cloud infrastructures and the tangibility of internet power’](#)

‘dirty’ sources.<sup>16</sup> Some estimates suggest that ‘a single desktop computer requires 240 kilograms of fossil fuels, 22 kilograms of chemicals, and 1,500 kilograms of water to manufacture’.<sup>17</sup>

In response to these recognitions of its harms, different concepts and terminology have emerged to present new or alternative ways of thinking about how data should be conceptualised and used. We describe these collections of ideas and terms as narratives. It is our hypothesis that these narratives have occurred as a direct response to the identification of the problems and scandals we’ve highlighted above. Many of these narratives have an explicitly ‘normative’ element, meaning they set out how data *should* be used.

**Data stewardship** is one of the narratives that we encountered in our research. For the [Ada Lovelace Institute](#), data stewardship is a ‘responsible, rights-preserving and participatory concept [which] aims to unlock the economic and societal value of data, while upholding the rights of individuals and communities to participate in decisions relating to its collection, management and use’.<sup>18</sup> The [Royal Society](#) uses data stewardship to describe a body mandated to ensure responsible use of data,<sup>19</sup> and the [Mozilla Foundation](#) uses it as a term to describe the act of empowering agents in relation to their own data and guidance toward a societal goal.<sup>20 21</sup> The [Aapti Institute](#) describes data stewardship as a ‘paradigm which explores how the societal value of data can be unlocked while considering what it takes to empower individuals/communities to better negotiate on their data rights’.<sup>22</sup> [The GovLab](#)’s focus on data stewardship is slightly different. It seeks to professionalise the role of data stewards, who are ‘agents of change in an organisation, responsible for determining what, when, how and with whom to share private data for public good’.<sup>23</sup>

The concept of stewardship inherently involves a dynamic relationship between at least two parties; stewarding data relates to the role of ‘looking after it on behalf of others’. According to the [Cambridge Dictionary](#), [stewardship](#) ‘of something is the way in which that person controls or organises it’. The idea of stewardship has been around for a long time, and prior to its application to data, has tended to focus on the control or organisation of companies, land and money. For example, [the National Trust](#) was established in 1895 to steward the ‘Nation of lands and

---

<sup>16</sup> Monserrate, S. G. (2022), ‘[The Cloud Is Material: On the Environmental Impacts of Computation and Data Storage](#)’

<sup>17</sup> *ibid*

<sup>18</sup> Ada Lovelace Institute (2021), ‘[Participatory data stewardship](#)’

<sup>19</sup> The Royal Society (2017), ‘[Data management and use: governance in the 21st century](#)’

<sup>20</sup> Mozilla Foundation (2020), ‘[Data Stewardship - What is it and why does it matter?](#)’

<sup>21</sup> Mozilla Foundation (2020), ‘[What Does it Mean? | Shifting Power Through Data Governance](#)’

<sup>22</sup> Aapti Institute, Mohamed, S. (2022), ‘[Situating Civil Society Organisations on the Stewardship Spectrum](#)’

<sup>23</sup> Young, A. (2018), ‘[About the Data Stewards Network](#)’

tenements (including buildings) of beauty or historic interest' for the United Kingdom. Applied to data, the role of looking after something could be done on behalf of individuals, communities, companies. The concept of 'data stewardship' embeds the idea of *having to negotiate relationships when using data*.

Another widely recognised way of thinking about how data should be used is **data ethics**. At the ODI, we define 'data ethics' as:

*A branch of ethics that evaluates data practices with the potential to adversely impact on people and society – in data collection, sharing and use.*<sup>24</sup>

We encountered two broad discussions and sets of approaches within the data ethics discourse. First, there are those that seek to develop and apply principles to shape how data systems should function – such as the UK Government's [Data Ethics Framework](#).<sup>25</sup> A global review of AI ethics principles was undertaken by Jobin, Ienca, and Vayena,<sup>26</sup> who reviewed 84 documents on AI ethics. They undertook a content analysis of these documents and found nine main, high-level principles: Transparency, Justice & fairness, Non-maleficence, Responsibility, Beneficence, Freedom & autonomy, Trust, Sustainability, and Dignity.

Other approaches suggest that, to understand if something is ethical, a *process of negotiation of ethical considerations* is needed, between all the actors in a particular context.<sup>27</sup> Although principles were seen by our interviewees as being a useful starting point for engaging with data ethics, we found a desire to move towards this more process-based approach. For instance, we spoke to the London Office for Technology and Innovation, which had [developed a service to support data ethics](#) discussions:

“ ... the objective of [the service] is to actually provide meaningful hands-on help to [organisations] around data ethics... helping them discover via learning what good practice looks like and develop their organisational capabilities in local government in London.

---

<sup>24</sup> The Open Data Institute (2021). [The Data Ethics Canvas](#)

<sup>25</sup> GOV.UK, Central Digital & Data Office, (2020), '[Data Ethics Framework](#)'

<sup>26</sup> Nature Machine Intelligence, Jobin, A, Ienca, M, & Vayena, E, (2019), '[The global landscape of AI ethics guidelines](#)'

<sup>27</sup> Minds and Machines, Morley, J, Elhalal, A, Garcia, F, Kinsey, L, Mökander, J, & Floridi, L, (2021), '[Ethics as a Service: A Pragmatic Operationalisation of AI Ethics](#)'

In our research, we also came across numerous references to [data for good](#).

In 2022, Data.org published a report by Jake Porway, co-founder of DataKind, which included a taxonomy of over 600 initiatives that described themselves as either ‘data for good’ or ‘AI for good’.<sup>28</sup> They identified different ways in which organisations in this ecosystem talk about the idea of ‘for good’.<sup>29</sup> One way is in the context of for-profit data products, where ‘data for good’ focuses on reducing the harms that result from making and misusing these products.

Another use focuses on the types of uses that data products are put to, where ‘data for good’ involves supporting people to use data more effectively in civil or charitable contexts.

In 2018, Involve, Understanding Patient Data and the Carnegie UK Trust undertook an engagement process with members of the UK public to understand how they thought of [public benefit](#) in relation to the sharing of health data.<sup>30</sup> They identified three themes that were central to the idea of data having wider public benefit:

- providing intrinsic benefits to society through access to better public services
- delivering improved outcomes for communities
- enabling research

Their research reiterated the importance of the wider public being a beneficiary from the use of data and those benefits being vital for obtaining social licence for collecting, using and sharing data.

[Connected by Data](#) has set out a similar narrative around [effective or good data governance](#).<sup>31</sup> This brings together the broader concept of data governance, which it defines as making decisions about data, with Connected by Data’s commitment to social justice. The key features of good data governance here is that it is collective, democratic, participatory and deliberative.

Related to this, we encountered the concept of [human-centred data](#), which places the needs of people at the centre of designing data strategy and systems. The [MyData](#) community emerged to support organisations

---

<sup>28</sup> data.org, (2022), ‘[A Taxonomy for AI / Data for Good](#)’

<sup>29</sup> data.org, Porway, J, (2021), ‘[Charting the ‘Data for Good’ Landscape: Principles and Methodology](#)’

<sup>30</sup> Involve, Scott, K, (2018), ‘[Data for Public Benefit: Balancing the risks and benefits of data sharing](#)’

<sup>31</sup> Connected by Data, Tennison, J, (2022), ‘[Effective Data Governance](#)’



implementing human-centred data in practical ways, primarily via a code of conduct<sup>32</sup> and guiding principles that include concepts like transparency, agility and diversity.<sup>33</sup> [The World Economic Forum](#) similarly utilises a human-centred approach to data in its work.<sup>34</sup>

**Data justice** is a way of thinking about data that takes into account broader issues of equity, value and power. Work by the Alan Turing Institute on Advancing Data Justice produced a number of practice guides, research outputs and policy practice pilots on data justice, including an integrated literature review<sup>35</sup> that identified six pillars of data justice: power, equity, access, identity, participation, and knowledge.

We identified two understandings of the idea of ‘data justice’ in our interviews. Firstly, there was one which most closely resembled underlying principles – ie ‘justice and fairness’. Here, data justice was seen as being achieved when the value produced by data, typically financial value, is distributed ‘fairly’ between organisations collecting data and data subjects. As one of our interviews stated:

“ *Data justice is ... related to value creation and the distribution of value.*

Others use an alternative understanding of ‘data justice’, particularly those concerned with critical data studies or non-Western approaches. In this context, data justice is defined not only as a fairer and more equitable distribution of the proceeds of data use, but also as making concrete steps towards people having a much greater say in the use of data. There can be particular concern for marginalised groups or those who have been disproportionately affected by technologies. Under this approach, those groups would have the power to design the ways data about them, or that impacts them, is collected and processed. For example, the organisation [Data for Black Lives](#) (D4BL) is ‘a movement of activists, organisers, and scientists committed to the mission of using data to create concrete and measurable change in the lives of Black people’.<sup>36</sup> In addition, for some of our interviewees, data justice incorporated the idea that data should be used to rectify and rebalance inequalities of power between different groups and across geographies:

---

<sup>32</sup> MyData (n.d.), ‘[Code of conduct](#)’

<sup>33</sup> MyData (n.d.), ‘[Guiding Principles](#)’

<sup>34</sup> World Economic Forum, Bettinger, K, (2021), ‘[12 ways a human-centric approach to data can improve the world](#)’

<sup>35</sup> Leslie, D, Katell, M, Aitken, M, Singh, J, Briggs, M, Powell, R, Rincón, C, Chengeta, T, Birhane, A, Perini, A, Jayadeva, S, & Mazumder, A, (2022), ‘[Advancing Data Justice Research and Practice: An Integrated Literature Review](#)’

<sup>36</sup> Data for Black Lives (n.d.), ‘[Data for Black Lives](#)’



*...for example, I'm in the United States, making sure that Black and Indigenous people are able to develop their own standards about the data that gets collected on them or that they're doing the data collection... data justice takes it beyond responsible data to something that maybe looks a little bit like reparations even.*

The narrative of **data sovereignty** can also be seen as empowering individuals, organisations and nations to self-determine how, when and under what conditions data can be used.<sup>37 38</sup> **Indigenous data sovereignty** is a related idea and is defined as 'an assertion of the rights and interests of Indigenous Peoples in relation to data about them, their territories, and their ways of life'.<sup>39</sup> The CARE Principles for Indigenous Data Governance have the 'primary goals of (1) fostering Indigenous self-determination by enhancing Indigenous use of data for Indigenous pursuits and (2) honoring the 'FAIR Guiding Principles for scientific data management and stewardship' (Findable, Accessible, Interoperable, Reusable), while ensuring data sharing on Indigenous terms. The hope is that these CARE principles will be adopted by the research and data community so that Indigenous data is used appropriately in these contexts.

We came across the concept of **data feminism** in our research. According to D'Ignazio and Klein, it is 'a way of thinking about data, data systems, and data science that is informed by the rich history of feminist activism and feminist critical thought'.<sup>40</sup> Their work developed into a list of principles and a book<sup>41</sup> exploring them in more detail. The seven principles of data feminism are: 1) Examine power, 2) Challenge power, 3) Elevate emotion and embodiment, 4) Rethink binaries and hierarchies, 5) Embrace pluralism, 6) Consider context and 7) Make labour visible.

Some of these narratives are more tightly defined than others, leading to more specific recommendations and approaches. For instance, with 'data ethics', there is still a need to identify an underlying ethical approach or

---

<sup>37</sup> Datasphere Initiative, De La Chapelle, B, & Porciuncula, L, (2021). '[We Need to Talk About Data: Framing the Debate Around Free Flow of Data and Data Sovereignty](#)'

<sup>38</sup> International Data Spaces (n.d.), '[Data Sovereignty](#)'

<sup>39</sup> Data Science Journal, Carroll, S. R, Garba, I, Figueroa-Rodríguez, O. L, Holbrook, J, Lovett, R, Materechera, S, Parsons, M, Raseroka, K, Rodriguez-Lonebear, D, Rowe, R, Sara, R, Walker, J. D, Anderson, J, & Hudson, M, (2020), '[The CARE Principles for Indigenous Data Governance](#)'

<sup>40</sup> Responsible Data, D'Ignazio, C, & Klein, L. F, (2021), '[The Seven Principles of Data Feminism](#)'

<sup>41</sup> D'Ignazio, C, & Klein, L. F, (2020), '[Data Feminism](#)'

framework that would be applied to understand whether something is 'ethical' – e.g. bioethics, human rights, utilitarianism. 'Data ethics' is probably best understood as describing a range of approaches or a field of endeavour. In contrast, narratives such as Indigenous data sovereignty and data feminism are much more specific in their suggestions for the 'right' way to use data and approaches to take to operationalise it.

# Chapter 2:

## Responsibility

The concept of responsibility is prominent among the field of alternative narratives for data collection, use and sharing, and indeed technology more broadly. As a general definition, responsibility refers to an individual or group's accountability for their actions or decisions and the impact they have on others.<sup>42</sup>

It also requires individuals and groups to consider the consequences of their decisions and to act with integrity and ethical standards. It is visible in a wide range of real-world situations and sectors. We see the term used to talk about [investments](#), [fashion](#), [consumption and production](#), all of which have the common ambition of minimising the negative impacts – largely social and environmental – of a given activity. The concept of responsibility is one that many people use in their day-to-day lives to describe themselves or others.

In the world of data and technology, **the language of responsibility is used by lots of different actors**. Oxfam and other organisations from the humanitarian sector have put both policy and practical resources together to help organisations manage humanitarian data responsibly.<sup>43</sup>

Philanthropic organisations, such as the Omidyar Network, use the concept of responsibility to fund and advocate for healthier digital ecosystems.<sup>44</sup> From finance companies like Mastercard<sup>45</sup> to public broadcasters like the BBC,<sup>46</sup> organisations are talking about responsible data or responsible AI principles.

But the proliferation of uses also highlights one of our research key findings: that **there is no consistent definition of responsibility in regards to data**. Instead, there are clusters of ideas around what it entails. For example:

- The [Responsible Data](#) community defines responsible data as (1) ensuring people's rights to consent, privacy, security, and ownership, (2) protecting information processes, including collection, analysis, storage, presentation, and reuse of data; and (3) respecting values of transparency and openness.

---

<sup>42</sup> The Britannica Dictionary, (n.d.), '[Responsibility Definition](#)'

<sup>43</sup> Oxfam (2017), '[Responsible Data Management training pack](#)'

<sup>44</sup> Omidyar Network, (n.d.), '[Responsible Technology](#)'

<sup>45</sup> Mastercard, Monti, J, (2019), '[Mastercard Establishes Principles for Data Responsibility](#)'

<sup>46</sup> BBC, Macgregor, M, (2021), '[Responsible AI at the BBC: Our Machine Learning Engine Principles](#)'

- [The United Nations Office for the Coordination of Humanitarian Affairs \(OCHA\)](#), defines data responsibility as ‘the safe, ethical and effective management of personal and non-personal data for operational response, in accordance with established frameworks for personal data protection’.<sup>47</sup>
- The [Internet Society](#) defines responsible data handling as a means of applying ethical principles of transparency, fairness and respect to how we treat the data that affects people’s lives.<sup>48</sup>

We also came across some variations of what the concept of responsibility was applied to – such as ‘Responsible and ethical innovation’, ‘Responsible data/AI’, ‘responsibly unlocking the value of data’,<sup>49</sup> and ‘acting responsibly with data’.

## ResponsibleData.io

[Responsibledata.io](#), a community facilitated by The Engine Room, has been promoting the concept of ‘responsible data’ since 2014. In 2022, the [Responsible Data Forum](#) is celebrating its [8th birthday](#) and has attracted over 1,000 members from a broad set of countries and sectors. This community has played an important role in promoting the concept of responsible data and many NGOs and other organisations refer to the community’s understanding of ‘responsible data’.

The concept was defined collectively by the ResponsibleData community (see above) and is tied to their perception of social justice, which is not solely focussed on the redistribution of resources, but also on the recognition and redistribution of existing power imbalances too (data is creating power imbalance, and responsible data would mean correcting these inequalities). One of our interviewees said:

“ *Responsible data is about making sure that you're taking into account the unintended or unexpected things that can happen as a result of working with data. So is making sure that people's rights to consent to privacy, to ownership of the data are taken into account when doing social justice work or social change work, but also being transparent and*

<sup>47</sup> United Nations Office for the Coordination of Humanitarian Affairs (2021), ‘[Data Responsibility Guidelines](#)’

<sup>48</sup> Internet Society (2019), ‘[Policy Brief: Principles for Responsible Data Handling](#)’

<sup>49</sup> Datasphere Initiative, De La Chapelle, B, & Porciuncula, L, (2022). ‘[Hello Datasphere - Towards a systems approach to data governance](#)’

*open about what you are doing with that data.  
[...] For me, when I think about responsible data, it's inevitable to also think about power dynamics that come with working with data.*

In addition to there being no consistent definition of responsibility, **some use the term 'responsible' in relation to data interchangeably with terms such as 'trustworthy' and 'ethical'**. One of our interviewees said, 'I think being responsible means being ethical. It's the same thing.' On the other hand, **some see responsibility as broader and more expansive than other concepts**, making ethics a concern that a responsible approach has to take into account, among others. According to another interviewee we spoke to,

“ *responsibility is a reflective process [...] where the researcher or innovator [...] will have to anticipate ethical problems, legal, social, societal concerns.*

A number of our interviewees see the term **'responsibility' as an accessible way to think about data practices**. The interviewees spoke about how concepts like data ethics can require a high level of knowledge or understanding to engage with, whereas the idea of 'responsibility' is used by many people on a regular basis well beyond conversations about data. This commonsense dimension of the term 'responsibility' was seen as a benefit – the fact that many people understand it could lead to its ideas being more readily applied.

For some, **being responsible with data means adhering to a set of principles; for others it's a process**. Similarly, in the world of data ethics, there is a tendency to describe responsibility as a set of principles to urge people to use data more responsibly. On the contrary, some of our interviewees found this principles-based approach to be inferior to a more process-based one, and that being more responsible with data should be an ongoing and constantly renegotiated effort. One of our interviewees stated:

“ *I think it incorporates more than the set of principles that we have specifically associated with data [...] it's a commitment to those principles that's anticipatory, reflective, reflexive, inclusive, deliberative, and responsive to emerging challenges.*

According to our interviewees, one of the main issues with the principles-based approach is the translation and operationalisation of responsibility into actual data practices. Responsibility is not solely a linear tick-boxing exercise; rather, it is ‘an integrated and iterative set of processes with the necessary capacities to support them’.<sup>50</sup> A further issue with a principles-based approach, common in existing frameworks and guidance around responsible data, is that some of the principles are not very well elaborated, grounded or tailored to the sector they are embedded in.

“*Operationalising that in practice is another thing and there’s trade offs around performance. [...] I think the frameworks are great. They give quick orientation, they provide a framework for working, but I think there’s still real challenges in that translational space between what does that mean in practice and how you solve those problems.*”

Our interviewees from the humanitarian sector described taking a very ‘rules-based’ approach to responsible data practices. The pressing and critical situations in which the humanitarian sector tends to operate, and the sensitivity of the data it works with, especially related to vulnerable populations, means that organisations have been encouraged to set ‘very strong guidelines’ to ensure a minimum viable level of responsibility is achieved:

“*We’re managing data about highly vulnerable people who are facing one of the most difficult moments of their lives, [...] there’s a lot of responsibility on humanitarian organisations to not expose these people to additional risk when managing their data.*”

**Associations with ‘responsibility’ differ depending on geographical and cultural context.** The term ‘being relative’ was a common theme in our research, meaning that the idea of responsibility depends on the relevant context and can change over time. There was a strong feeling among our interviewees that responsibility is not understood to mean the same thing around the world.

---

<sup>50</sup> United Nations Office for the Coordination of Humanitarian Affairs (2016), ‘[Building data responsibility into humanitarian action](#)’

One of our interviewees stated:

“ And that is why sometimes when I talk about responsible data governance or responsible stewardship, we have to understand people’s individual differences or socio-cultural values in which we are talking about responsibility. We have to consider people’s expectations within that culture. We have to understand people’s fears and concerns. These are unique fears and concerns. The fears in Europe might not be the fear in Africa.

Some of our interviewees felt that **the use of concepts such as responsibility could help to address the limited, Western viewpoint that skews most of the discourse around data** and digital technologies.

<sup>51</sup> That is of course, only if these concepts acknowledge structural inequalities and power struggles in regard to data, as well as involving and empowering historically marginalised data subjects. One of our interviewees said:

“ There is a huge unequal distribution of the respective benefits of the data economy at the current stage. And that any future vision of the digital society should be oriented towards not only not aggravating the inequalities but potentially reducing them in many regards, particularly when we are talking about using data for achieving sustainable development goals or other aspects.

---

<sup>51</sup> Eke, D. O. et al, (2023), ‘Responsible AI in Africa, Social and Cultural Studies of Robots and AI’



Several of our interviewees said **being responsible with data involves going above and beyond data protection and other laws**, which was seen as the bare minimum of a responsible approach to data. As one interviewee suggested:

“ I think privacy is important, but there are so many more issues that don't get quite as much coverage even related to privacy [...] I mean like even the framing of privacy over the years has, has been very sort of obviously heavily influenced by GDPR but has taken a very individual approach like individual privacy rather than thinking of collective notions of things like privacy.

Some interviewees pointed out that many countries lack robust legislative frameworks for data, and therefore going above and beyond a low or non-existent bar was imperative to acting responsibly.

Similarly, one interviewee explained how a more responsible approach to data should encapsulate the positive consequences of collecting, using and sharing data, and the importance of creating a 'responsible data culture'<sup>52</sup> within an organisation:

“ We chose the term responsibility because it is more encompassing than data protection or privacy. It goes beyond just the management of personal data and beyond limiting data sharing or restricting data management, and emphasises the importance of sharing data safely, ethically and effectively where appropriate.

Lastly, we found that **the language of responsibility has been adopted by large technology companies**. Responsible data, AI or innovation teams have appeared at [Meta](#) and Salesforce,<sup>53</sup> and Google has developed a set of [Responsible Data Practices](#), perhaps in response to the

---

<sup>52</sup> United Nations Office for the Coordination of Humanitarian Affairs (2021), 'Data Responsibility Guidelines'

<sup>53</sup> Salesforce (2020), 'How Salesforce Is Building a Culture of Responsible Technology – and Why it Matters'

increasing demand for more transparent and responsible innovation from civil society and governments.<sup>54</sup> This language is often also used in the context of technologies such as AI, where there is interest in ‘developing AI systems that will not only be compliant to laws (including human rights provisions) but that are socially/culturally sensitive and acceptable as well as be ethically responsible’.<sup>55</sup> This has been evident among the recent progress and discourse around large scale language models (LLM) and generative AI, where models like Bloom have sought to differentiate themselves from other LLMs<sup>56</sup> through an ‘ethical charter’ that sets out how the model seeks to be inclusive, diverse, reproducible and open.<sup>57</sup> Some of our interviewees were wary of ‘responsibility-washing’ in this context, however.

---

<sup>54</sup> The Economist (2013), ‘[The coming tech-lash](#)’

<sup>55</sup> Eke, D. O. et al, (2023), ‘[Responsible AI in Africa, Social and Cultural Studies of Robots and AI](#)’

<sup>56</sup> The Guardian, Helmore, E, (2023), ‘[We are a little bit scared](#)’: OpenAI CEO warns of risks of artificial intelligence’

<sup>57</sup> Big Science (n.d.), ‘[BigScience Ethical Charter](#)’

# Chapter 3:

## Responsible data practices and interventions

As well as understanding how the concept of responsibility is used, we sought to explore how it shapes real-world data practices. By data practices, we mean behaviours, policies and processes related to collecting, using and sharing data.

In some instances, our interviewees found it difficult to translate the theoretical ideal of responsibility into concrete practice and identify examples:

“ *... like how do you operationalize a big concept like diversity? There isn't a single definition... it means different things in different contexts, it changes over time.*

Some interviewees felt that this wasn't only difficult in the context of 'responsibility' but that it's an issue that arises often when working on 'data ethics' and other concepts discussed in this report. Interviewees reflected on the effort that it took to anticipate issues and to explore mitigations:

“ *it's hard to always anticipate what the impact or implications are gonna be on different groups. We don't always have the domain expertise, we don't always have the experience ourselves. We'll be speaking from... our own position. So... it's due diligence... that we are creating space ... to understand what the impacts and implications might be in different contexts for different groups at all times.*

However, others were able to describe responsible data practices, which we've arranged into the following typology, according to how they arose from the research interviews:

- Practices that seek to protect privacy
- Practices that seek to reduce bias in data processing
- Practices that seek to be transparent or participatory
- Practices that seek to be reflective
- Practices that seen to deliver a public good
- Practices that seek to increase equality and equity
- Practices that seek to reduce data collection to when it's necessary/proportionate
- Practices that seek to reduce negative impacts from the use of data

## Practices that seek to protect privacy

We encountered practices that considered both technological arrangements - such as the design of technology architecture or the use of particular technologies - as well as legal or governance protections. For instance, one interviewee highlighted the development of privacy legislation, such as GDPR in Europe, as a responsible practice. Another interviewee talked about different tools, such as statistical methods that help reduce the risk of identity disclosure, as well as an approach for classifying data and information types into different sensitivity levels in a given response context, e.g. the humanitarian response in Ukraine.<sup>58</sup> Other practices that were raised to us as being 'responsible' included having clear data management policies. For instance, the [OCHA Data Responsibility Guidelines](#) recommend actions such as an information-sharing protocol and creating a data registry.

## Practices that seek to reduce bias in data processing

Responsible data practices were also seen as being those that act to reduce biases in the processing of data which might disadvantage certain groups in experiencing positive outcomes or mean they would be more likely to experience negative outcomes. For instance, the NHS undertook detailed assessments of the representativeness of a chest imaging dataset that was collected during the Covid-19 pandemic, and developed

---

<sup>58</sup> United Nations Office for the Coordination of Humanitarian Affairs, (2022), '[Data and Information Sensitivity Classification Ukraine](#)'

standards and guidance for developers and innovators using this dataset for AI training purposes. Salesforce<sup>59</sup> has also built flags into its products, highlighting certain attributes such as ZIP code to users as potentially contributing to bias.

## Practices that seek to be transparent or participatory

Other examples of responsible data practices could be grouped into those where there was a commitment to a form of user-centred design or activities to ensure the participation of data subjects or people affected by data use. Some examples include:

- [An IoT project](#) delivered by a number of South London Boroughs, which produced a map of the locations of all of the public sensors installed by the project.
- [Camden council's Data Charter](#),<sup>60</sup> which was created by local residents and is intended to guide how the council's data is used.
- The City of Helsinki has developed a 'human-centred approach to data'<sup>61</sup> in its design of services and has applied this to the use of citizen data in its provision of proactive public sector services.

Many of our interviewees considered participation to be an important part of the implementation of responsible data practices.

“ I think a big part of responsible data and I think one of the trickiest parts, is to make sure that the people represented in the data or the people whose data belong to have a say in what happens with that data. It's not necessarily an easy thing to incorporate participatory approaches to the work, but a big part of how I see responsible data.

---

<sup>59</sup> World Economic Forum, (2022), 'Responsible Use of Technology: The Salesforce Case Study—WHITE PAPER'

<sup>60</sup> Camden Council, (2022), 'Camden publishes Data Charter promising safer and ethical use of data'

<sup>61</sup> World Economic Forum, Bettinger, K, Ziskind Ferrari, J, Lähteenoja, V, (2021), 'Empowered Data Societies: A Human-Centric Approach to Data Relationships'

## Practices that are reflective

A number of interviewees highlighted the same example of an organisation responding to a data breach as a particularly good example of a responsible practice. In particular, the International Committee of the Red Cross's (ICRC) response to a cyber attack,<sup>62</sup> and Oxfam instituting a moratorium on using biometrics data while it consulted on its Biometric and Foundational Identity Policy were recognised for their openness.<sup>63</sup> In both of these examples, it seemed that the organisations involved were reflective about these issues, which interviewees saw as being 'responsible'.

Conversely, there was a group of practices that were seen as irresponsible, which were characterised by a lack of reflection and consideration. For instance, we heard the example of the data breach related to the United Nations High Commissioner for Refugees (UNHCR) collecting data on Rohingya refugees, which was passed to the Bangladesh government and subsequently shared with the government of Myanmar. This sharing of data means that the 'same military that conducted the (most recent) genocide against the Rohingya now holds the biometric data of the population it has tried to eradicate'.<sup>64</sup> Interviewees also made reference to data being used inappropriately in the banking sector without consent. We also heard that in Uganda, citizen's personal data has been breached as a result of the government recycling sensitive paper documents into paper bags used in retail contexts.

Another set of practices that are reflective are those that involve teams coming together over a period to explore particular issues or go through a set process for assessment. This was seen as being characteristic of the work undertaken by the City of Helsinki in its adoption of a human-centric approach to data.<sup>65</sup>

## Practices that seek to deliver a public good

A number of interviewees highlighted practices that were considered responsible because they were in the service of delivering a form of public good or public benefit, such as improving health, education, research, or environmental outcomes. Some of the examples of data being used for the public good raised in the interviews include:

---

<sup>62</sup> International Committee of the Red Cross, (2022), '[Sophisticated cyber-attack targets Red Cross Red Crescent data on 500,000 people](#)'

<sup>63</sup> Views & Voices, Eaton-Lee, J. (2021), '[Oxfam's new policy on biometrics explores safe and responsible data practice](#)'

<sup>64</sup> The New Humanitarian, Rahman, Z, (2021), '[The UN's refugee data shame](#)'

<sup>65</sup> World Economic Forum, Bettinger, K, Ziskind Ferrari, J, Lähteenoja, V, (2021), '[Empowered Data Societies: A Human-Centric Approach to Data Relationships](#)'

- A manufacturer of medical devices bringing anonymised data from its customers to train AI learning algorithms to detect tumours.
- Using drone footage to identify pathways through agricultural land for nomadic cattle herders in Africa, thereby reducing tension with farmers.
- The [Fair Tech Collective](#)'s programme of citizen science air quality monitoring, where data is shared with the local public health authority.

These practices highlight the importance of considering a wide range of different publics (or groups of people) that could either benefit or be affected by the use of data. The common denominator of these different practices is that they go beyond focusing only on avoiding harms with data, and proactively show how data can be used to tackle some of society's biggest challenges.

## Practices that seek to increase equity

There was also a selection of practices highlighted as responsible because they sought to redress broader societal inequities or recognised that data can perpetuate them. For instance, an interviewee highlighted the work undertaken by Salesforce in its organisational commitment to ethical and humane use of technology. One of the changes it made was to replace the use of the term 'master' in its naming conventions.<sup>66</sup> The use of this term was seen as problematic as it was 'a historically insensitive term associated with slavery'. The New Zealand-based organisation Local Contexts has developed a hub that empowers Indigenous groups to label data about cultural material with community-specific conditions regarding access and use.<sup>67</sup>

## Practices that seek to reduce data collection to when it's necessary/proportionate

Some interviewees also raised that practices of data minimisation, meaning that data collection should take place only when necessary, and the principle of proportionality in regards to data, were key topics of data responsibility.<sup>68</sup> According to them, the sector has focused too much in the past decade on collecting data that they did not even have time to analyse, which was not considered as responsible practice.<sup>69</sup>

<sup>66</sup> World Economic Forum, Green, B, Ratté, E, Smallwood, R, Zhang, C, (2022), 'Responsible Use of Technology: The Salesforce Case Study—WHITE PAPER'

<sup>67</sup> Local Contexts, (n.d.), 'Traditional Knowledge Labels'

<sup>68</sup> CartONG, (2022), 'Responsible data management toolbox, Q&A section'

<sup>69</sup> CartONG, 'Information Management Portal, Learning corner'

## Practices that seek to reduce the negative impacts of the use of data

Many of the practices we learnt of seek to avoid negative, or unintended, impacts of data. Within these practices, there was a recognition that those who experience negative outcomes do not always have the power to change the outcomes themselves.



*We are capable of mapping pretty much every local community for one... third of the rainforest .... we're starting the conversations like, do we make this public? And... we haven't decided yet. ... we're gonna be thinking about this very, very thoroughly and carefully ... because then it's open season on these communities.*

In addition to asking interviewees to identify responsible data practices, we asked about interventions they were familiar with that had been used to encourage, support or otherwise drive these practices. Although there was a recognition that interventions to support responsible practices were still relatively nascent, our research surfaced a number of interventions that different actors were employing to stimulate more responsible practice.

Many of our interviewees described **responsible data frameworks** that set out how to work responsibly with data. These included the [Oxfam Responsible Program Data Policy](#) (2015), [RD 101: Responsible Data Principles](#) (2018), and [The OCHA Data Responsibility Guidelines](#) (2021). This approach is so well established that in 2018 the US-based Center for Democracy and Technology published the report 'Responsible Data Frameworks In Their Own Words',<sup>70</sup> which sought to compare a range of these publications, predominantly in the humanitarian data sector. More practical resources have been developed, such as the [IFRC data playbook](#), The Engine Room [RAD tipsheets](#), CartONG's [Information Management Resource portal](#) and [Responsible data management toolbox](#), and practical resources put together by [MERL tech](#).

As we discuss in Section 1, there is often a gap between principles and practise for normative approaches to data. As some of our interviewees flagged, many of these current frameworks do not help people to translate from principles to practice.<sup>71</sup>

---

<sup>70</sup> Center for Democracy and Technology, (2018), '[Responsible Data Frameworks: In Their Own Words](#)'

<sup>71</sup> Minds and Machines, Morley, J, Elhalal, A, Garcia, F, Kinsey, L, Mökander, J, & Floridi, L, (2021), '[Ethics as a Service: A Pragmatic Operationalisation of AI Ethics](#)'



A number of interviewees talked about using [technology](#) to support responsible data practices.

“ We are also [funding] a data collaborative platform... so that when these people know how to collect and manage the data, they can be able to share it and collaborate with each other on that.

One interviewee working in the health field mentioned a number of different technological interventions they were putting into place to support the more responsible use of data. For instance, they are seeking to develop a chatbot to deliver services to excluded groups and are exploring whether some users would like to have more say over data and about them using a personal data store. Another example given was the use of privacy enhancing technologies, such as [federated learning](#), to facilitate greater sharing of sensitive data in ways that ensure data protection and privacy.

We also heard a desire from our interviewees for [independent training, advice or certification](#) to support the adoption of responsible data practices:

“ Can I have like a bunch of one-on-ones where I explain to you the problem and, and together going through the checklist... you sort of walk us through whatever issue.

Many felt there was an opportunity for non-profit organisations like the ODI, or others whose work is described in this report, to help others implement responsible data practices. This ongoing support was seen as an important addition to the frameworks mentioned above, enabling organisations to more effectively translate some of these ideas around responsibility into practices and actions in their contexts. In our literature review we also encountered a number of commercial organisations advertising services to support organisations to use data and AI more responsibly.

Several of our interviewees mentioned that training would be a suitable intervention to support organisations to use data responsibly. However, there was a recognition by some that training needs to be specific enough to be useful in a particular context:

“ I think training and upskilling is good... But a lot of the training is quite generic training in this area and I think there is a need in certain sectors and industries and contexts for more bespoke tailored training to types of issues... that puts those issues in, in the context in which people are grappling with them.

One of our interviewees described an appetite from businesses for a certification scheme for the responsible use of data. While acknowledging the difficulty in doing this, they said that:

“ ...certification schemes come up all the time as ideas of what we can do. The core issue that we face here is the auditing of it ... unless we are getting resourced to do this ... it becomes very tricky. So we haven't solved this as an organisation. It does seem like there's lots of appetite from people. That this is an approach that they would be interested in.

Interviewees discussed the role of **government intervention** in driving responsible data use. Legislative interventions such as the [California Consumer Privacy Act \(CCPA\)](#) were often highlighted as drivers of responsible practices. However, it was also raised that legislation and regulation in this area was difficult and that for governments, not regulating in this area could be seen as a competitive advantage:

“ So how do we ensure responsibility? The UK's approach is almost like a middle ground... in that we're currently in this kind of light touch regulation space where ... we're not regulating, but the work...that's happening across other organisations in the UK government is about... how... we work with businesses.

Policymakers and funders have also intervened by **funding and introducing new organisations** to shape data practices. For example, in the UK, the government established the [Centre for Data Ethics and Innovation](#) to support the trustworthy use of data and AI. Similarly, the

Nuffield Foundation launched the Ada Lovelace Institute in the wake of the Cambridge Analytica scandal with a remit to ‘anticipate ethical questions raised by emerging technologies and their application’.<sup>72</sup>

Many other philanthropic donors and NGOs have launched research and funding agendas with a focus on ‘responsibility’. Some funders, such as the [Data Futures Lab at the Mozilla Foundation](#), focus on supporting organisations collecting and managing data, while the Responsible Technology focus from the Omidyar Network includes supporting a wider ecosystem of actors to realise the positive benefits of data and technology.<sup>73</sup>

In some contexts, **financial incentives or controls** were seen as a possible way to support more responsible ways of using data. Particularly in large organisations or governments, one interviewee suggested that there could be governance processes to determine whether particular data projects would get funding, based on a commitment to, or proof of, responsible data use. This could operate in a manner similar to ‘spend controls’ operated in the UK central government in relation to the implementation of the [Government Digital Service Service Standard](#).

A significant intervention described to us was the **fostering of communities of practice and alliances between organisations**.

“ we are trying to have a network... We are looking at all organisations in the data governance space in Africa (and) want to bring them together and share what they have, share the opportunities. So we don't want it to be like a library ... (we) want it to be more engaging, ...don't just push what you published, engage and share opportunities with each other.

In a similar vein, interviewees talked about developing a network of champions within and across organisations who would work to identify and support the development of responsible practice.

“ The only thing that works is individual champions... if you get... a handful of individual champions who support each other working in different organisations, that's... where you get the real leverage in our sector...

---

<sup>72</sup> The Nuffield Foundation, (2018), ‘[The Nuffield Foundation announces new £5 million Ada Lovelace Institute to examine profound ethical and social issues arising from the use of data, algorithms and AI](#)’

<sup>73</sup> Omidyar Network (n.d.), ‘Responsible Technology’

The Responsible Data community facilitated by The Engine Room is a good example of this type of approach, as well as the Data Stewards Network setup by The GovLab, which seeks to ‘connect responsible data leaders from the private and public sectors seeking new ways to create public value through cross-sector data collaboration’.<sup>74</sup>

---

<sup>74</sup> GovLab (n.d.), ‘[Data Stewards Network](#)’

# Chapter 4:

## What we mean by responsible data stewardship

We've used the concept of data stewardship in our work at the ODI, generally describing it simply as 'the collection, maintenance and sharing of data'.

In our work on [data institutions](#) in recent years, we've expanded on this by suggesting that these organisations are stewarding data 'on behalf of others towards public, educational or charitable aims'. We've learnt much about what this involves in practice through our work with [INSIGHT](#) to design and run a participatory data governance process, and with dozens more data institutions through our stimulus funding, peer learning networks and mentorship activities.

However, at the outset of this research, our understanding of what makes for good data stewardship was largely anecdotal and based mainly on our experiences with specific use cases. We were driven to develop and articulate an interpretation of *responsible data stewardship* to:

- add a normative element to our description of data stewardship, based on the ideas we explore in Section 1.
- provide a more critical lens that we can use in our work to help others design and practice data collection, use and sharing.
- offer an interpretation that other organisations might find useful to consider or adopt in their work.
- more easily identify how our work is related – or not – to other organisations using similar language.
- reduce the opportunity for the concept of data stewardship to be misappropriated.

Going forward, we understand responsible data stewardship to be:

*an **iterative, systemic** process of ensuring that data is collected, used and shared for **public benefit**, mitigating the ways that data can produce **harm**, and addressing how it can **redress structural inequalities**.*

## Iterative

To us, responsible data stewardship is a negotiated and reflective process. Because contexts vary and change over time, mitigations and approaches to collecting, maintaining and sharing data need to constantly evolve. Responsible data stewardship is far from a box-ticking exercise but rather an effortful, proactive process, including consistent and meaningful engagement with examples of responsible practice from a variety of contexts. Implementing ongoing, reflective processes requires organisations to take this journey seriously, which means investing staff time and financial support into working through these challenges, institutionalising and embedding practices. Success should be judged by the activities and outcomes, including the ability to anticipate and address irresponsible practices.

## Systemic

The impacts of data collection and use are rarely fully within the control of any one organisation. Organisations need to develop a systemic view of their data practices that links how choices made around data have impacts outside of the organisation. This requires an understanding of the data ecosystem in which it operates, as well as a realistic and critical appreciation of the interests, capacities, power and vulnerabilities of actors elsewhere in the ecosystem. Internally, this means thinking about how data is approached, not just from a technical perspective, but also considering the governance, allocation of resources and the values of those around it. Responsible data stewardship will be the result of decisions made throughout an organisation and will require collaboration between teams.

## Public benefit

Stewarding data responsibly involves ensuring it's used and shared for the benefit of others, rather than only for the benefit of the organisation that holds it. This means proactively exploring how data can be used for positive impact, rather than only focusing on avoiding its potential harms. Organisations should seek opportunities to use or share data to tackle our

biggest challenges, such as the climate crisis, risks to our public health, improving education opportunities worldwide, and more. Organisations should meaningfully involve and empower data subjects and other stakeholders in this process of realising the wider benefits of data.

## Harm

Alongside seeking positive impact from the use of data, responsible data stewardship involves identifying and reducing harmful impacts to individuals and communities. These harms can arise in different ways – a lack of security or effective governance might mean that sensitive data gets into the hands of bad actors, or the products that data is used to create could have harmful unintended consequences for their users. Importantly, mitigating the potential harms of data involves going above and beyond legal requirements around privacy, security and transparency.

## Redress structural inequalities

The collection, use and sharing of data always occurs within a wider system of relationships, value exchanges and power imbalances. These have real-world consequences for data. It may lead to it being approached purely as a resource to be extracted from people for commercial gain, or to embedding harmful bias in the way we see the world into the data we collect and use to develop new technologies. Responsible data stewardship may involve meaningful new communication with data subjects and other stakeholders, or adopting alternative forms of governance. In some cases, it may involve stopping or not collecting data at all.

# Chapter 5:

## What's next?

First, we are keen for feedback on this work. In particular, we're interested in any emerging normative concepts and language we may have missed in our research, efforts to practice responsible data stewardship in the real world and views on our interpretation of responsible data stewardship.

We recognise that there is an appetite for more practical guidance and support in this area. Our research has highlighted that, in isolation, high-level principles can have limited impact and can be difficult to apply. In our next phase of work on responsible data stewardship, we will explore the potential interventions that we could take to support organisations to operationalise the learnings of this research. We intend to put particular emphasis on concepts such as 'public benefit', 'reducing harm' and 'redressing structural inequalities', exploring what's needed to practice them in different organisational, geographic and sectoral contexts.

This research has also reiterated to us the importance of narratives as a way of shaping how organisations think of and behave with data. In our next phase, we intend to refine and share this narrative with others, especially outside of our typical networks and conversations. We're hopeful about the potential for this – and related the ideas we discuss in this report – to help bring people and organisations together around a vision for [a world where data works for everyone](#).



# Endnotes

## Methodology

This research began with a review of literature related to new or alternative approaches to data governance. In identifying the literature to include in this review, we prioritised recent practical resources (handbooks, playbooks, etc) and research papers, as well as reports from think tanks, NGOs and governmental organisations.

Alongside this desk research, we conducted two internal workshops, two workshops with a stakeholder group of experts from the field of data governance and 18 interviews with experts and practitioners working with data from across a variety of sectors and geographies. A quarter of the interviewees (4/18) were based in the Global South with the rest based in the UK, North America or Europe. 12 were working for third sector organisations, four were working for public sector organisations and two worked for private companies.

The interviews were semi-structured and broadly covered a discussion about the organisation's work and focus, what people understood as responsibility and how this compared to terms such as 'data ethics' or 'data justice', and then moved on to discuss specific data practices and interventions that could support their further use. All interviews were transcribed and then coded and thematically analysed using the software Dovetail.

# Acknowledgements

This report was written by Joe Massey, Sasha Moriniere and Ed Parkes (Emerging Field), with supporting contributions from Yusuff Adigun, Jane Crowe, and Jack Hardinges. We would like to thank the experts that we interviewed for their participation in this research:

- Damien Eke, De Montfort University
- Phillip Ayazika, Pollicy
- Bonaventure Saturday, Pollicy
- Rhia Jones, BBC
- Sam Roberts, Department for Science, Innovation & Technology
- Maeve de France, CartONG
- Jules Caron, Project Canopy
- Sam Nutt, London Office of Technology and Innovation
- Kimmy Bettinger, World Economic Forum
- Brhmie Balaram, NHS AI Lab
- Stefano Wroblewski, InfoAmazonia
- Bertrand de la Chapelle, The Datasphere Initiative
- Eliana Quiroz, Hivos
- Paul Currion, Humanitarian Consultant
- Sara Baker, Engine Room
- Barbara Paes, Engine Room
- Emelia Williams, Open Environmental Data
- Janis Wong, The Alan Turing Institute
- Jos Berens, The United Nations Office for the Coordination of Humanitarian Affairs
- Ed Teather, AWS

We'd also like to thank the members of our stakeholder group, provided strategic input to the direction of this project:

- Lara Groves, Ada Lovelace Institute
- Valentina Pavel, Ada Lovelace Institute
- Tim Davies, Connected by Data
- Maximilian Gahntz, Mozilla
- Vinay Narayan, Aapti Institute
- Soujanya Sridharan, Aapti Institute
- Nikita Japra, Patrick J. McGovern Foundation
- Katie Shaw, Open Supply Hub
- Aiden Peppin, formerly Ada Lovelace Institute
- Kira Allman, formerly Ada Lovelace Institute
- Phillipa Nazari, Greater Manchester Combined Authority and Transport for Greater Manchester

Finally, we'd like to thank the Patrick J. McGovern Foundation for funding this research.