Chapter One: What is (in) Good Data?

**Monique Mann, S. Kate Devitt, Angela Daly**

# Introduction: Why Good Data?

In recent years, there has been an exponential increase in the collection, aggregation and automated analysis of information by government and private actors. In response to this there has been significant critique regarding what could be termed ‘bad’ data practices in the globalised digital economy. These include the mass gathering of data about individuals, in opaque, unethical and at times illegal ways, and the increased use of that data in unaccountable and discriminatory forms of algorithmic decision-making.

This edited collection has emerged from our frustration and depression over the previous years of our academic and activist careers critiquing these dystopian ‘Bad Data’ practices. Rather, in this text on ‘Good Data’ we seek to move our work from critique to imagining and articulating a more optimistic vision of the datafied future. We see many previous considerations of Bad Data practices, including our own, as only providing critiques rather than engaging constructively with a new vision of how digital technologies and data can be used productively and justly to further social, economic, cultural and political goals. The objective of the Good Data project is to start a multi-disciplinary and multi-stakeholder conversation around promoting good and ethical data practices and initiatives, towards a fair and just digital economy and society. In doing so, we combine expertise from various disciplines and sectors, including law, criminology, justice, public health, data science, digital media, and philosophy. The contributors to this text also have expertise in areas such as renewable energy, sociology, social media, digital humanities, and political science. There are many fields of knowledge that need to come together to build the Good Data future. This project has also brought together academic, government and industry experts along with rights advocates and activists to examine and propose initiatives that seeks to promote and embed social justice, due process rights, autonomy, freedom from discrimination and environmental sustainability principles.

We acknowledge that we are not the first people to start thinking about ethical data and data justice.[[1]](#footnote-1) But we view ‘Good Data’ as being a broader, and more open-ended, idea than data ethics or data justice, which may confine the conversation, for instance, to philosophical questions or a consideration of ‘inserting’ ethics in technical aspects, and not engage with wider political, historical, social, cultural, technological and economic issues.

We also wanted to take a more global approach to Good Data given much of the discussion and critique on data practices emanates from the Global North/West, in the spirit of creating and supporting Southern scholarship about data issues.[[2]](#footnote-2) In this edited text there are contributions from five continents which we view as a step towards broadening the good and ethical data discussions out from the Global North/West, although we acknowledge our position of privilege as academics based in the ‘Global North-in-South’.[[3]](#footnote-3) Furthermore, we acknowledge limitations of the book in this regard that we do not have a contribution from the African continent, and also our choice of language in the form of English. We hope that in the future Good Data discussions will be broadened to include more contributions from authors based in more geographical locations and in more languages than we have managed with this book.

The Good Data project was developed and initiated by us when we were all based together at Queensland University of Technology (QUT) in Brisbane/Meanjin - located on traditional Turbal land in what is now known as Australia - in late 2017. Each of us had been working on research engaging with social science aspects of data and digital technologies, and Angela and Monique had also been very active in digital rights activism in Australia. The situation in Australia was then, and still is, far from ‘best practice’ in data and digital issues - the lack of an enforceable constitutional right to privacy, the Australian government’s ongoing digital colonialism perpetuated against Indigenous peoples, refugees and other marginalised people and a myriad of other ways in which unethical data practices were being implemented.[[4]](#footnote-4) We had spent a lot of time and energy criticising these practices from both academic and activist perspectives, but we realised that we had not presented a positive alternative: how could data and digital technologies be designed and used in ‘good ways’, for ‘good’ purposes? If digitisation and data are inevitabilities, then we have to (re)imagine the kind of digitised world and data we want to see rather than only offering a naysaying critique of the status quo.

The Good Data project formally commenced with a multi-stakeholder workshop hosted by us and funded by the QUT Faculty of Law in late 2017. We designed the workshop to gather representatives of different academic disciplines and people who had in some way created, used or implemented ‘Good Data’ practices. The workshop was invite-only, and we organised an outreach public event in the evening featuring well-known digital rights activist Asher Wolf (@Asher\_Wolf) in conversation. We would like to thank Thoughtworks Brisbane for hosting our public event and providing catering for the audience.

We wanted the workshop and public event to be just the beginning of our enquiry into Good Data. Given the interest and engagement in our project, we thought that the next step was a book project on Good Data with an open call for contributions. From the beginning we knew that the book would have to encompass and showcase Good Data practices itself. Firstly, we are delighted to be working with the Institute of Network Cultures (INC) given their commitment to quick open access publishing on innovative topics related to digital network cultures. So many texts related to Good Data reside behind paywalls and as a result are not widely accessible, particularly to those outside of the academy. Furthermore, with the increasing and all-encompassing datafication of society and the economy, we were keen to issue this collection on Good Data in a rapid and timely manner (whose publication is taking place a little over a year after our initial Good Data workshop). We have extensive experience ourselves in waiting for our (academic) writing to see the light of day (often behind a paywall) and so we also appreciated the speed with which the INC could facilitate this collection getting out into the world. We also asked contributors for short chapters which would be accessible to non-specialists in order to widen the book’s appeal.

Perhaps the first question that should be asked when writing a book about Good Data is what is the nature of ‘good’. What is ‘goodness’? In our call for papers we were deliberately agnostic with regards to a conceptual analysis of ‘good’ because it intentionally sought transdisciplinary, culturally diverse and inclusive contributions. Foundational questions on ‘goodness’ for society and individuals from a western philosophical perspective - a perspective in which we ourselves are situated - might consider ‘goodness’ as increasing wellbeing (including hedonic, desire-theories and objective list-based theories), sustainability, fairness, justice, virtue and so on. For example, how would a utilitarian or Rawlsian design a smart city? How should app developers incorporate feminist ethics of care - that prioritises relationships and responsibilities over the individual - into their choice architecture? Yet, data discourses from underrepresented, disenfranchised and disempowered voices need to be prioritised rather than hegemonic conceptual structures. For example, (how) can autonomous vehicle data regulation incorporate intersectional feminist or Marxist political agendas? When and who should participate in radical data resistance and erasure? We believe this book is just one step into a long term project of interrogating diverse ethical, cultural and political theoretical frameworks into data practices.

Since we view ‘Good Data’ as a discussion which transcends atomised academic fields, we are pleased to see contributors and contributions to this book coming from cross/multi/transdisciplinary perspectives. Another of our aims was to move the discussion on Good Data beyond one disciplinary or professional sphere, and we are also pleased to see academics, activists, civil society representatives and technical experts contribute chapters to this book. Finally, we gave authors the option to have their chapter peer-reviewed or editor-reviewed. We thought that the option of editor review would ensure that people from other fields beyond academia could contribute to the collection.

In the next section we offer an overview of the authors’ contributions to this collection under each theme. We acknowledge that many of the contributions are relevant to more than one theme, but we have grouped them as best we can under some headings which give a flavour of the chapters:

1. Good Data Manifestos and Practices;
2. Good Data and Justice;
3. Good Data as Open and Shared Data;
4. Good Data Activism and Research; and
5. Good and Smart Data.

Finally, we would like to offer our sincere thanks to our research and editorial assistants who have helped us bring this book to fruition. Our particular thanks go to Dr Kayleigh Murphy and Ms Harley Williamson without whose hard work and dedication we would not have been able to complete the book project within such a speedy timeframe. We would also like to thank Anna Carlson who assisted us at the beginning of the project and organised a wonderful outdoor public event in November 2017 on ‘Data Activism and Digital Rights’[[5]](#footnote-5) (at which Angela spoke) for the Brisbane Free University, ‘an autonomous space in which the empowering processes of teaching and learning belong to everybody’.[[6]](#footnote-6) Anna also wrote a series of Good Data blogposts for the DATACTIVE Big Data Sur blog.[[7]](#footnote-7) Last, but far from least, we would like to thank all of the peer reviewers who contributed their time and expert insights to commenting on the peer-reviewed chapters, thereby strengthening the final versions of the chapters, and this book overall.

# 2. What’s in the Good Data book?

#### 

**Theme 1: Good Data Manifestos and Practices**

We kick off the book with a selection of manifestos and guidance on what ‘Good Data’ is, or should be, and how it should be used, or not.

First, in **Chapter 2**, we are delighted to have a multi-authored contribution on Good Data Practices for Indigenous Data Sovereignty and Governance from a group of prominent international Indigenous scholars, namely Raymond Lovett, Vanessa Lee, Tahu Kukutai, Donna Cormack, Stephanie Carroll Rainie and Jennifer Walker. Indigenous Data Sovereignty (IDS) and Indigenous Data Governance are Indigenous-led movements and practices through which Indigenous peoples are setting their own visions for good data regarding data generated and collected by and about them. IDS movements and practices can be seen as a manifestation of Indigenous peoples’ sovereignty more generally and as an alternative vision of data, centreing Indigenous peoples’ rights to self-determination and autonomy.[[8]](#footnote-8) IDS also challenges conventional, western colonial data practices, which have been utilised against Indigenous peoples since colonisation and continue to be used against them in the digital environment. The authors set out the context for, and emergence of, IDS movements and provide an overview of IDS developments including the IDS networks such as Te Mana Raraunga, the Māori Data Sovereignty Network in Aotearoa/New Zealand.

In **Chapter 3**, we then move to Claire Trenham and Adam Steer’s Good Data Manifesto, which draws on their practical experience as data scientists working with geospatial data. Their manifesto sets out a series of ‘Good Data’ questions that data producers and consumers should ask, constituting a set of principles which can be used to guide data collection, storage, and re-use. According to the authors, good data should be: usable and fit for purpose; collected with respect to humans and their rights, and the natural world; published; revisable; and form useful social capital. They draw on various examples to illustrate these questions and principles, with a focus on geospatial data which is often voluminous, ubiquitous and also - significant from a data protection perspective - personal.

**Chapter 4** by Miren Gutiérrez, considers the question of ‘good enough data’, particularly for social activism. ‘Good enough’ is contrasted with institutional, government or corporate data collection that may be systematic, but also imbued with control mechanisms to protect data stakeholders. Good enough data is thus a way to promote the use of data by grass roots activists and citizens to impose political pressure for social ends. The author thus defends ‘good enough data’ as data created, valued and interpreted by ordinary people, such as environmental data and citizen sensing. Good enough to sustain ongoing legal investigations. She offers the example of the Berkeley Human Rights Investigation Lab (HRC Lab) that used the Syrian archive to categorise chemical attacks on a Syrian city as evidence for violations of international humanitarian law as well as violations other regulations and treaties.

Finally, we turn to our next manifesto in **Chapter 5**, this time for good energy data, authored by Declan Kuch, Naomi Stringer, Luke Marshall, Sharon Young, Mike Roberts, Iain MacGill, Anna Bruce and Rob Passey. The authors are Australia-based energy researchers who view a close link between access to energy data and the country’s transition to a sustainable and just community-based energy future, which they argue is currently hampered by some major incumbent energy sector businesses and politicians. Rooftop solar (PV) panels are popular additions to Australian homes but individuals do not have access to the data about the energy they produce and consume. Access to this data would empower individuals and collectives such as community energy groups, and accordingly could hasten Australia’s take-up and implementation of sustainable energy in a sustainable, communal way. The authors provide a series of recommended actions in their manifesto which would lead to this goal.

**Theme 2: Good Data and Justice**

Data justice is a term which has become very prominent in recent times. We acknowledge in particular the work of the Data Justice Lab based at Cardiff University,[[9]](#footnote-9) and their conceptualisation of ‘data justice’ as ‘a conceptual foundation for exploring how data-driven surveillance implicates different understandings of social justice as well as a potential action-building tool for addressing such implications’.[[10]](#footnote-10) We also acknowledge Taylor’s work on elucidating a concept of international data justice based on three ‘pillars’ of ‘(in)visibility, (dis)engagement with technology and antidiscrimination’.[[11]](#footnote-11)

The chapters in this theme contribute to, and extend, the idea of data justice, through case studies on data justice topics in different areas such as criminal justice, genomics and cross-border data flows. We see these chapters as also contributing to discussions on digital criminology,[[12]](#footnote-12) by widening the discipline’s traditional focus beyond ‘The Cyber’ and cybercrime to look at wider socio-political contexts of digital citizenship.

What is meant by ‘fairness’ is a central consideration in ‘Good Data.’ This question is addressed by McNamara, Graham, Broad, and Ong on (racial) bias[[13]](#footnote-13) in **Chapter 6** that examines actuarial models of criminal justice. McNamara and colleagues examine assumptions that underpin recidivism prediction methods in Australia, with the objective of identifying and rectifying bias, specifically in relation to domestic violence cases. Significantly, the authors draw attention to the politics of data collection and statistical inference. For example, they question Indigenous status being selected as a predictor variable, and argue that the social context of algorithmic decision-making is an important determinant of outcomes in criminal justice. Their examination of the predictive validity of risk assessment tools demonstrates that there are serious consequences for ‘trade-offs’ in adopting various definitions of ‘fairness’. Accordingly, the authors re-design the predictive model in order to reduce the discriminatory impact of the model towards Indigenous persons, yet this comes at the ‘trade-off’ of reduced predictive accuracy. With these findings identified, the authors outline various approaches to algorithmic fairness in all stages of data collection and processing. Their analysis demonstrates the importance of incorporating explicit fairness criterion into predictive models - and making trade-offs in fairness explicit - as Good Data practice.

Ozalp’s chapter - **Chapter 7** - presents a detailed case study of the ByLock case concerning unlawful data access leading to mass persecution of dissidents in Turkey. Through the lens of moral panic analysis Ozalp recounts a case study of what can go bad without good information security, providing a concrete real world example of the oppressive potential of bad data practices, and while questioning what we can learn for Good Data. In doing so, Ozalp outlines how digital communication technologies and strong information security are essential to support what he terms a ‘good democracy’ – including protection for the rights such as freedom of expression, political thought, religion, association and privacy. Accordingly, it is argued that counter-surveillance practices, online anonymity and encryption tools are integral to a good democracy.

Next under the data justice theme is **Chapter 8:** Arnold and Bonython’s examination of genomic data. They argue that recent technological developments in this area, a ‘perfect storm’ may be brewing where governments want to implement population-wide genomic databases, private corporations attempt to commodify genomic data through the intellectual property system, and direct-to-consumer genomic testing opens a Pandora’s Box of legal, political and ethical issues. Their chapter is a timely and crucial contribution to conceptualising ‘Good Data’ in this area. Underpinning ‘goodness’ regarding genomic data is, according to the authors, a fundamental respect for human dignity which ought to manifest, for instance, in truly consensual, fair and transparent data collection and use. The authors also emphasise an ‘ethic of responsibility’ regarding genomic data which ought to be implemented in various ways, including through regulation and government intervention, professional codes, public education, and decision-making by public bodies funding research and development in this area.

The final contribution to this section is **Chapter 9**, where Melashchenko tackles the contentious issue of data localisation,[[14]](#footnote-14) or stipulations that data be physically held on machines within the territory of a particular country or jurisdiction. Data localisation policies in some countries evidence the materiality of data. Data does not have an ethereal existence but exists in a physical location. This is significant as principles of territoriality underpin the state’s legitimacy to police and govern. The issue of data localisation goes to the heart of (digital) geopolitics and legal geographies of digital technologies, since it may or may not be desirable from different perspective for data to be held in a particular country, or for it to be held ‘offshore’ in order to evade certain laws and policies. Melashchenko considers data localisation in the context of trade, data justice and privacy. He identifies that data localisation policies are far from a monolith and may differ in their intensity and detail, and accordingly maps out some ‘variables’ and ‘types’ of these policies. This is followed by the introduction of some normative criteria against which data localisation policies can be assessed as being ‘smart data regulation’ which facilitates data justice, and ultimately a vision of Good Data.

**Theme 3: Good Data as Open and Shared Data**

Our next theme centres ideas of open data and shared data as forms of Good Data. Discussions of open data have preceded much of the contemporary focus on privacy as being a main ‘Good Data’ issue, although concerns about the unintended consequences of (some) open data for privacy and other rights are now prominent. Reconciling tensions between open data and data protection (and other interests) is a key challenge for this area.

In the first chapter in this theme - **Chapter 10** - Gray and Lämmerhirt consider the social life of the Open Data Index, a civil society project that aims to measure and influence how much government data is made available. In particular they attend to how the index organises participation and data politics, comparing indexes to the political mobilisiation afforded by rallies, petitions and hashtags. Indexes represent social facts but also reshape the social world - quantifying and thus enabling normative claims about data practices of different countries, and encouraging participation to resolve them. The Open Data Index aims to intervene around what is considered ‘Good Data’ by assessing the extent to which the publication of government data conforms with specific legal and technical norms and conventions. As a database about data, it can measure accountability and intervene on official regimes of quantification and datification. This is relevant to other chapters considered in the volume including smart cities and good enough data that consider the role of government versus citizen data and the role of data activism. The discussion of the role of indexes (ratings and rankings) is also relevant to the data visualization chapter that acknowledges the impact of visualisations on the epistemology of users. Visualizations are heavily employed by indexes to attempt political and social change (e.g. Corruption Perceptions Index). The Open Data index chapter is thus able to contribute to a larger conversation around the importance of empowering citizens with good data.

We then turn to **Chapter 11**, where Porlezza addresses open source ethics in data journalism, broadly speaking the use of computer science and statistical analytic methods in journalism practice, including programming, data analysis and visualisations to inform reporting. The author identifies four open source normative concepts to guide activity in this area, namely transparency, participation, tinkering, and iteration, which he argues can facilitate Good Data journalism. Drawing on empirical research conducted with data journalists in Italy and Switzerland, actual data journalism practices are assessed against these four concepts to evaluate the extent to which these activities fulfil and align with open source ethics. Various differences are identified in how data journalists in both countries approach, adhere to and implement these ethical principles in their work, which Porlezza mainly attributes to structural differences between the journalism environment in the two countries rather than individual journalists’ own moral codes.

In the final chapter in this theme**, Chapter 12**, Ho and Chuang critique neoliberal data protection models which emphasise individual autonomy and choice through concepts such as consent and anonymisation. Instead, the authors propose an alternative model for data use and sharing based on community participation in decision-making and self-governance, drawing from commoning scholarship and practice, and computational methods. They look to examples including data cooperatives to analyse how community participation can occur regarding data governance, in ways which can facilitate use and sharing of that data but are also trusted by the collective. In this way, communal data sharing models present a Good Data alternative to the current widespread prorprietary and extractive models.

**Theme 4: Good Data Activism and Research**

Ethical research and activism is a key component of ‘Good Data’. In an age of fighting back against Bad Data, government surveillance and corporate capture of civil society and academia (see for instance Funding Matters’ recent resistance to Palantir’s sponsorship of the Amsterdam Privacy Conference),[[15]](#footnote-15) it is crucially important for the Good Data agenda to outline approaches to ethical activism and research. Chapters within this theme advance and consider what ‘Good’ activism and research looks like and outline principled approaches to conducting Good Data activism and research. Chapters within this theme also consider the interaction between activists and academics research agendas and outline models for ethical data-activist collaborations.

The theme’s first chapter - **Chapter 13** - by Zeffiro investigates research ethics boards at funding bodies and universities in Canada and found that they share a piecemeal approach to research ethics in the face of changing technologies. She proposes a united effort to create ethical guidance for using social media data for whole of life-cycle research that acknowledges the diversity of needs of different interdisciplinary researchers. At its core researchers must be transparent about their methodologies for generating, collecting, processing analysing and storing social media data. Social media data is likely to be harvested without informed consent, without concern for the welfare of participants and potentially without sensitivity to vulnerable or at risk participants. Zeffiro notes that third party disclaimers on social platforms are not sufficient for ethical and informed consent by research participants. Participant data accrued from social media participants must be kept anonymous, yet researchers must acknowledge that confidentiality cannot be guaranteed for data sent via the internet by any third parties. Researchers must recognise that their social media dataset has been generated by human participants who are likely unaware of how their inputs have been quantified and that they are complicit with the platforms themselves if they rely on them for research data. Zeffiro argues that the term ‘c/overt research’ should be used by researchers to acknowledge the ethical challenges with collecting and using social media data and the limitations of research ethics boards. Zeffiro proposes researchers question who they feel accountable towards, are self-reflective with regards to their own situated perspective and identify their duties to participants; maintaining a flexible and adaptable approach, rather than seeking a ‘one-size-fits-all’ solution to research.

Writing reflectively from the experiences of the DATACTIVE team at the University of Amsterdam, in **Chapter 14** Kazansky, Torres, van der Velden, Wissenbach and Milan consider what data activism research for the social good could look like aligned with the Good Data agenda. They question what forms of active and productive roles academics can play in advancing data activism, and also research agendas. In doing so, they examine the co-production of knowledge and the specific work of the DATACTIVE team and their direct involvement as activists with the communities that they study. Their main contribution is the proposal of an ‘engaged’ approach and epistemology to research that aims to contribute to activist causes – that is, doing research ‘with’ rather than ‘about’. They outline an approach to ethical data activist research as a process rather than a checklist and as inspired by Association of Internet Researchers (AoIR) ethical codes and feminist ethics of care.

Similarly, in **Chapter 15** Poletti and Gray outline what is ‘Good’ when it comes to critical digital studies, and advance an emancipatory method inspired by Marxist and critical ethical approaches. They argue that ‘Good Data’ is that which can be used to critique power dynamics associated with the use of data, and with a focus on economic and technological environments and contexts in which they are generated. They commence their examination by drawing attention to global informational capitalism and the asymmetric and extractive exploitation of data by companies such as Google, Facebook, Amazon etc, while advocating for reflection on the production system where data are created and collected and sold – and the tensions between ethical research and capitalist research (such as, for example, as revealed by the Facebook-Cambridge Analytica scandal). Poletti and Gray highlight the challenges to critical researchers in producing valid and ethical research in a data ecosystem of capitalist production and also that is formed and exists in cross/multi jurisdictional contexts. Drawing from the work of Antonio Casilli, Christian Fuchs and Karen Gregory they propose an approach to critical ethical data research that considers the economic and political order, and ground data ethics in a critique of neoliberal economic systems and digital labor and capitalist relations. They conclude that Good Data are data that can be used to bring about change in modern informational capitalism, but for this to occur there is a need to challenge the dominant rhetoric and further reflexivity in critical digital studies.

In **Chapter 16** Wieringa, van Geenen, van Es and van Nuss focus on a particular kind of research data communication format: network visualisations. Network visualisations are used to represent the geometric interconnectedness and interrelatedness of data, providing a more nuanced way of experiencing complex data than normal charts, graphs or tables. However, network visualisations can bias readers into believing the data presented is objective and complete; rather than interpretive and limited. The authors argue that the assumptions, methodologies and justifications behind the visualisation need to be more transparent and have created a plug-in for common data visualisation tool ‘Gephi’ to make them more accountable. Specifically their fieldnotes plugin allows users to export the details of the ethnographic, working process, including iterations of the graph file over time. The authors argue that the plugin is of relevance to critical data scholars more widely.

**Theme 5: Good and Smart Data**

The book’s final theme focuses on the changes ubiquitous interconnectedness brings to our cities, homes, personal and interpersonal information ecosystems. Governments, research institutions and corporations are invested in an innovation agenda that relies on extensive access to citizen data via smart phones; urban and domestic surveillance and the Internet of Things (IoT) to create ‘smart’ algorithms for ‘smart cities’ and ‘smart homes’. Families, groups and communities share personal data in homes and online and have collective interests beyond those of the individual. Technologies are usually touted as bringing convenience, efficiency, ease and wellbeing improvements to consumers, often in overtly technological determinist terms. However, ethical, regulatory and legal frameworks often lag behind consumer trust in these devices and networks. This theme brings together authors who consider data activism and citizen protection under the onslaught of data in private and public spheres. The authors consider citizen use of public, private and interpersonal data, offering insights and good data practices to protect individuals and groups.

In the first chapter in this theme, **Chapter 17**, Ritsema van Eck argues that mapmaking has changed as ‘smart cities’ gather data via video, audio, and other kinds of Internet of Things (IoT) sensing devices. The data-streams they generate can be combined with volunteered data to create a vast multitude of interactive maps on which individuals are constantly (re)grouped on the basis of abnormality, deviation, and desirability. Instead of extending personal data protection rights to groups - that is awkward within the current European Union data protection framework, which is the chapter’s focus - the author suggests protection can be achieved via Data Protection Impact Assessments (DPIAs), which are mandatory to carry out when the ‘systematic monitoring of a publicly accessible area on a large scale’ necessary for mapmaking takes place. DPIAs can identify risks such as discrimination at an early stage. Furthermore, by including representatives of local (disadvantaged) groups, the strong performative qualities of maps can offer occasions for groups of citizens in smart cities to proactively shape the environments in which they live. However the author acknowledges that substantial legislative change would be required to the DPIA process to ensure affected data subjects and their representatives were included in the consultative process.

Smart cities are promoted as creating more economically thriving, social and environmentally sustainable cities. However, in **Chapter 18** Valencia and Restrepo argue that they are usually driven by governments and corporations that promote a neoliberal, colonialist agenda to retain power and influence over citizens through increased surveillance and data secrecy. Citizens’ data is collected, analysed and used to drive political agendas, often without citizen consent or access. The authors investigate the possibility of citizen led smart cities that could lead to data empowered citizens. Rather than rejecting datatification, the authors discuss citizen organisations and resistance communities that demand open data and the production of citizen-led data and software to produce bottom-up smart cities instead of top-down. They argue that true smart cities can only emerge from inclusive and citizen-led social data practices. They focus particularly on an instance of urban agriculture in Colombia, where a citizen-led environmental resistance movement endorses open data and software, citizen-data gathering, digital activism, community connection and communication. The chapter explores how data activism can progress an agroecological political agenda and social movements.

This is followed by **Chapter 19,** in which Bosua, Clark, Richardson and Webb investigate user control of personal data in a connected digital world. They argue that the Internet of Things has the paradoxical result of introducing a new era of ‘smart computing’ while reducing the intelligent control that individuals can exercise over their personal data. Users should be able to exert greater control over the collection, storage and use of their personal data. In this chapter the authors provide early design concepts of systems that could improve personal control of data including privacy, data protection and cybersecurity. They argue that personal data empowerment can be achieved through better design that make data flows more transparent to users. In particular they focus on IoT data that is particularly vulnerable because it is frequently unencrypted and uncontrolled. They propose creating an Intelligent Warning App DataMind, using ‘privacy-by-design’ principles to incorporate ‘nudges’ to alert individuals about data issues of potential interest to them, thereby empowering them to take control of their personal data.

In **Chapter 20,** the last chapter of the book, Flintham, Goulden, Price and Urquhart warn of a future in which the Internet of Things creates group data that overwhelm the efforts of individual group members to manage personal information that other members have access to - what they call ‘interpersonal data’, because it is drawn from, and carries consequences for, the relationships between intimate groups. The authors examine European Union law (as the most proactive regulatory regime in this space) using design fiction methodologies to consider what good interpersonal data might look like and how to avoid it becoming ‘bad data’ through inappropriate design, or legal consequence. Data in homes is often co-constructed, yet legal protection is constrained to individualised notions of one user, one device. Homes are shared spaces between diverse individuals who participate in wide range of social practices including access to and control of data. These homes comprise not just nuclear units and are not necessarily safe for all parties to have agency. The authors point to divorce, break ups and domestic violence as particular challenges, but also discuss the risks to the agency of teenagers and children. Thus the specific danger within homes it is not a distant bad actor who constitutes the greatest threat but those most intimately connected to individuals. Smart devices digitise domestic interpersonal relations demonstrating how IoT technologies carry novel implications for interpersonal relations, and the data generated around them. Designing smart devices with the law in mind goes some way towards good data practices, however users have latitude to change settings that open up new challenges within their context of use. Data driven technologies must respect interpersonal relationships, and the distribution of agency amongst them, both socially and legally. They must also, in doing so, recognise the moral choices they are making in involving themselves in these spaces, and redefining their possibilities. The next generation of smart devices should, potentially, actively and disruptively, deliver data protection norms into the home including informed and visible transactions around data and designing personalised privacy interventions. However, given limited legal protections and fast-paced technological innovation, it is possible that the best data at home is not smart at all.

# 3. What (else) is Good Data?

We are very pleased to include 20 Chapters of Good Data discussions with differing perspectives on the question of what Good Data is. But, like data itself, it is impossible for us to cover everything encompassed by ‘Good Data’ and accordingly we cannot offer a ‘complete’, ‘comprehensive’ or ‘perfect’ account of good data at this stage (if indeed ever). This book is but a partial reflection of the ‘Good Data’ reality or possibility. But that is exciting because we are just the beginning of a process of establishing Good Data frameworks, processes and guidelines.

As we mentioned earlier in this Introduction, we acknowledge that this book still has a bias to the Global North in the contributions we received and included in this book, and we also acknowledge that our choice of language also may have excluded contributions. For Good Data going forward, a more global approach needs to be taken to the issue, rather than just (re)centreing perspectives from the Global North, as already noted – and critiqued - by Arora.[[16]](#footnote-16) In order to see more Global South perspectives on Good Data, and many other topics, Global North(-in-South) scholars such as ourselves need to take more steps to be inclusive and facilitate such perspectives, for instance by providing resources such as translation, access to academic databases and other assistance to our colleagues and comrades in the Global South. This is a increasingly crucial as we see a more multi-polar digital world, with the rise of China as an Internet giant and the large online - and increasingly datafied - populations of countries such as India (with its highly controversial Aadhar whole-of-population biometric database), Indonesia and Nigeria - to name but a few.

In addition to more Global Southern perspectives on Good Data, we also think that more account should be taken of how Bad Data practices impact specifically and more acutely on marginalised people and communities, and that Good Data thinking and practices must taken an intersectional approach to acknowledging the different and intersecting oppressions faced by People of Colour, Indigenous people, queer people, and disabled people. We are very pleased and honoured to have the chapter from Lovett and others in this book on Indigenous Data Sovereignty. Going forward, Good Data conversations, including our own, must take more account of intersectional perspectives.

We would also like to see more scholarship and thinking at various levels on Good Data in the form of theories, methodologies and practices for Good Data, in order to gain power for individuals and communities. While we are critical of technological determinist views of data, digital technologies and datafication, we do not include much on technical aspects of data in this book. We do not think that ‘only’ ‘hard-coding’ rights, interests, values etc into technology is enough to ensure Good Data practices; but we also do not think that it is unimportant either. We would like to see more ideas, discussion and scholarship, from an interdisciplinary perspective, about how this might happen. Finally, we have included some domain or sector-specific contributions in this book on the question and idea of Good Data; but given that ‘Good Data’ might look somewhat different in different contexts, we would also like to see more analysis and recommendations for specific domains and sectors such as science, agriculture, health and education.

For us, Good Data also goes beyond the digital, including when studying the digital as researchers. Social media and other digital data are tempting for researchers as they offer rich, complex and extensive insights into human behaviours, attitudes, beliefs and desires. However, researchers utilising digital data sources such as social media must be self-reflective regarding their methodologies and acknowledge their complicity with platform ethical commitments (or lack thereof) when they benefit from harvested platform data. Researchers also need to appreciate the limits of data scraped from public sources that may not reveal a full picture of participant views but instead present a carefully filtered representation. So the fight is deeper than the methodologies themselves and indeed extends to challenging various aspects of the contemporary neoliberal culture of research as a whole. A call for Good Data for holistically ethical research encourages researchers to strive towards higher order ethical norms, beyond the minimum required of ‘do not harm’ and instead a cry to do more, to ‘do good’.

# 4. Next steps: How do we start building Good Data initiatives?

An important question following from this book is how do we start to build, develop and reward Good Data initiatives? This is the logical next step in the Good Data project: moving from theory and academic inquiry (as we present in this collection) to progressing initiatives in practice. Good Data is but the first step in a long journey towards a just or ethical digital economy and society. With the utmost humility, we admit that we do not have all the answers to this question, but we have some ideas as where to place the next steps, and we present them as open questions in this final section.

There is first and foremost a need to take stock and question what tools - conceptual, theoretical, practical, legal and technical - that we need to build the Good Data future. Despite being socio-legal researchers working within Law Faculties, we do not think that law and regulation is the solution,[[17]](#footnote-17) and there are limits to looking to the law (and data protection law).[[18]](#footnote-18) With that said, we do see some potentially promising provisions such as Article 25 of the EU’s General Data Protection Regulation on ‘data protection by design and default.’ But, we have questions about how this translates (if indeed if it is possible to do so) into the design or hardcoding of systems.[[19]](#footnote-19) We also recognise this also needs to be about more than privacy,[[20]](#footnote-20) and it may be more fruitful if we shift the focus away from privacy, and towards data ethics, human rights more generally and social justice as we have attempted to do here in this Good Data book.

There is also a critical question about how do we engage and empower technologists, designers and end-users in building Good Data initiatives and communities of change? How do we educate and encourage them to be more ethical in their work, and indeed protest Bad Data practices, as we have recently witnessed.[[21]](#footnote-21) How can we work better with technologists, designers and end-users to co-design and co-educate each other about Good Data ethics, ideas and practices?

It is here we see our future work heading. With *Good Data* we have moved our work from critique to imagining and articulating a more optimistic vision of the datafied future, with our enormous gratitude to this book’s contributors for assisting us in doing so. The next steps for us all are to build the Good Data future we want to see rather than letting governments and companies build a Bad Data future for us.

# References

Arora, Payal. ‘The Bottom of the Data Pyramid: Big Data and the Global South’, *International Journal of Communication* 10 (2016): 1681-1699.

*BFU presents: “Don’t be Evil” – Data Activism and Digital Rights.* Public Panel Discussion by Brisbane Free University, West End, Brisbane, 8 November 2017,https://brisbanefreeuniversity.org/2017/11/06/bfu-presents-dont-be-evil-data-activism-and-digital-rights/

Brisbane Free University. About, https://brisbanefreeuniversity.org/about/

Browne, Simone. *Dark Matters: On the Surveillance of Blackness*, Durham: Duke University Press, 2015.

Brownsword, Roger, Eloise Scotford, and Karen Yeung (eds). *The Oxford Handbook of Law Regulation and Technology*, Oxford: Oxford University Press, 2017.

Carlson, Anna. ‘Imagining ‘Good’ Data: Northern Utopias, Southern Lessons’, *Big Data from the South blog*, 25 May 2018 https://data-activism.net/2018/05/bigdatasur-blog-13-imagining-good-data-northern-utopias-southern-lessons/

Carrington, Kerry, Russell Hogg, John Scott and Maximo Sozzo. *The Palgrave Handbook of Criminology and the Global South*, Cham, Switzerland: Palgrave Macmillan, 2018.

Connell, Raewyn. *Southern Theory: The Global Dynamics of Knowledge in Social Science*, New South Wales: Allen & Unwin, 2007.

Data Justice Lab. https://datajusticelab.org/

Dencik, Lina, Arne Hintz and Jonathan Cable. ‘Towards data justice? The ambiguity of anti- surveillance resistance in political activism’, *Big Data & Society* 3.2 (2016): 1-12.

Ferguson, Andrew. *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, New York: NYU Press, 2017.

Funding Matters. https://fundingmatters.tech/

Goldsmith, Jack and Tim Wu, *Who Controls the Internet: Illusions of a Borderless World,* New York: Oxford University Press, 2006.

Hilderbrandt, Mireille. ‘Extraterritorial jurisdiction to enforce in cyberspace: Bodin, Schmitt, Grotius in cyberspace’, *University of Toronto Law Journal* 63 (2013), 196–224.

Koops, Bert-Jaap. ‘The trouble with European data protection law’, *International Data Privacy Law*, 4.4 (2014): 250-261.

Koops, Bert-Jaap and Ronald Leenes. ‘Privacy regulation cannot be hardcoded. A critical comment on the ‘privacy by design’ provisions in data-protection law’, *International Review of Law, Computers and Technology*, 28.2 (2014): 159-171.

Kukutai, Tahu and John Taylor (eds). *Indigenous Data Sovereignty: Towards an Agenda*. Canberra: ANU Press, 2016.

Lee, Murray. ‘Crime and the Cyber Periphery: Criminological Theory Beyond Time and Space.’ in Kerry Carrington, Russell Hogg, John Scott and Maximo Sozzo (eds), *The Palgrave Handbook of Criminology and the Global South*, Cham: Palgrave Macmillan, 2018, pp. 223-244.

Mann, Monique and Angela Daly. ‘(Big) Data and the North-*in*-South: Australia’s Informational Imperialism and Digital Colonialism’, *Television and New Media* (in press, 2019).

Mann, Monique and Ian Warren. ‘The digital and legal divide: Silk Road, transnational online policing and Southern criminology’, in Kerry Carrington, Russell Hogg, John Scott and Maximo Sozzo (eds.), *The Palgrave Handbook of Criminology and the Global South*, Cham: Palgrave Macmillan, 2018, pp. 245-260.

Mann, Monique, Ian Warren and Sally Kennedy. ‘The legal geographies of transnational cyber-prosecutions: Extradition, human rights and forum shifting’, *Global Crime* 9.2 (2018): 107-124.

Matzner, Tobias. ‘Why privacy is not enough privacy in the context of “ubiquitous computing” and “big data”’, *Journal of Information, Communication and Ethics in Society*, 12.2 (2014): 93–106.

Milan, Stefania and Emiliano Treré. ‘Big Data from the South: The beginning of a conversation we must have’, *DATACTVE Blog*, 16 October 2017, https://data-activism.net/2017/10/bigdatasur/

Montiero, Mike. ‘Ethics can’t be a Side Hustle’, *Dear Design Student,* 19 March 2017,https://deardesignstudent.com/ethics-cant-be-a-side-hustle-b9e78c090aee

Noble, Safiya Umoja. *Algorithms of oppression: How search engines reinforce racism*, New York: NYU Press, 2018.

Powell, Anastasia, Gregory Stratton & Robin Cameron. *Digital Criminology: Crime Justice in Digital Society*, New York: Routledge, 2018.

Salinas, Sara and Jillian D’Onfro, ‘Google Employees: We no Longer believe the company places values over profit’, *CNBC,* 27 November 2018, https://www.cnbc.com/2018/11/27/read-google-employees-open-letter-protesting-project-dragonfly.html

Sandvig, Christian, Kevin Hamilton, Karrie Karahalios and Cedric Langbort. ‘When the algorithm itself is a racist: Diagnosing ethical harm in the basic components of software’, *International Journal of Communication* 10 (2016): 4972-4990.

Shaban, Hamza. ‘Amazon Employees Demand Company Cut Ties with ICE’, *Washington Post*, 22 June 2018, https://www.washingtonpost.com/news/the-switch/wp/2018/06/22/amazon-employees-demand-company-cut-ties-with-ice/?utm\_term=.713db9c3092a

Taylor, Linnet. ‘What is data justice? The case for connecting digital rights and freedoms globally’, *Big Data & Society* 4.2 (2017): 1-14.

1. See Data Justice Lab, https://datajusticelab.org/ [↑](#footnote-ref-1)
2. See: Stefania Milan and Emiliano Treré, ‘Big Data from the South: The beginning of a conversation we must have’, *DATACTIVE Blog*, 16 October 2017, https://data-activism.net/2017/10/bigdatasur/; Kerry Carrington, Russell Hogg, John Scott and Maximo Sozzo, *The Palgrave Handbook of Criminology and the Global South*, Cham: Palgrave Macmillan, 2018; Raewyn Connell, *Southern Theory: The Global Dynamics of Knowledge in Social Science*, New South Wales: Allen & Unwin, 2007. [↑](#footnote-ref-2)
3. Monique Mann and Angela Daly, ‘(Big) Data and the North-*in*-South: Australia’s Informational Imperialism and Digital Colonialism’, *Television and New Media*, (in press, 2019). [↑](#footnote-ref-3)
4. Ibid. [↑](#footnote-ref-4)
5. *BFU presents: “Don’t be Evil” – Data Activism and Digital Rights.* Public Panel Discussion by Brisbane Free University, West End, Brisbane, 8 November 2017,

   https://brisbanefreeuniversity.org/2017/11/06/bfu-presents-dont-be-evil-data-activism-and-digital-rights/ [↑](#footnote-ref-5)
6. Brisbane Free University, About, https://brisbanefreeuniversity.org/about/ [↑](#footnote-ref-6)
7. Starting with: Anna Carlson, ‘Imagining ‘Good’ Data: Northern Utopias, Southern Lessons’, *Big Data from the South blog*, 25 May 2018,

   https://data-activism.net/2018/05/bigdatasur-blog-13-imagining-good-data-northern-utopias-southern-lessons/ [↑](#footnote-ref-7)
8. See also: Tahu Kukutai and John Taylor (eds), *Indigenous Data Sovereignty: Towards an Agenda*, Canberra: ANU Press, 2016. [↑](#footnote-ref-8)
9. Data Justice Lab, https://datajusticelab.org/ [↑](#footnote-ref-9)
10. Lina Dencik, Arne Hintz and Jonathan Cable, ‘Towards data justice? The ambiguity of anti-surveillance resistance in political activism’, *Big Data & Society* 3.2 (2016): 9. [↑](#footnote-ref-10)
11. Linnet Taylor, ‘What is data justice? The case for connecting digital rights and freedoms globally’, *Big Data & Society* 4.2 (2017): 8. [↑](#footnote-ref-11)
12. Anastasia Powell, Gregory Stratton and Robin Cameron, *Digital Criminology: Crime Justice in Digital Society*, New York: Routledge, 2018. [↑](#footnote-ref-12)
13. See also: Simone Browne, *Dark Matters: On the Surveillance of Blackness*, Durham: Duke University Press, 2015; Andrew Ferguson, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, New York: NYU Press, 2017; Safiya Umoja Noble, *Algorithms of oppression: How search engines reinforce racism*, New York: NYU Press, 2018; Christian Sandvig, Kevin Hamilton, Karrie Karahalios and Cedric Langbort, ‘When the algorithm itself is a racist: Diagnosing ethical harm in the basic components of software’ *International Journal of Communication*, 10 (2016): 4972-4990. [↑](#footnote-ref-13)
14. See also: Jack Goldsmith and Tim Wu, *Who Controls the Internet: Illusions of a Borderless World,* New York: Oxford University Press, 2006; Mireille Hilderbrandt, ‘Extraterritorial jurisdiction to enforce in cyberspace: Bodin, Schmitt, Grotius in cyberspace’, *University of Toronto Law Journal* 63 (2013), 196–224; Murray Lee, ‘Crime and the Cyber Periphery: Criminological Theory Beyond Time and Space’, in Kerry Carrington, Russell Hogg, John Scott and Maximo Sozzo (eds), *The Palgrave Handbook of Criminology and the Global South*, Cham: Palgrave Macmillan, 2018, pp. 223-244; Monique Mann, and Ian Warren, ‘The digital and legal divide: Silk Road, transnational online policing and Southern criminology’, in Kerry Carrington, Russell Hogg, John Scott and Maximo Sozzo (eds), *The Palgrave Handbook of Criminology and the Global South*, Cham, Switzerland: Palgrave Macmillan, 2018, pp. 245-260; Monique Mann, Ian Warren and Sally Kennedy, ‘The legal geographies of transnational cyber-prosecutions: Extradition, human rights and forum shifting’, *Global Crime* 9.2 (2018): 107-124. [↑](#footnote-ref-14)
15. Funding Matters, https://fundingmatters.tech/ [↑](#footnote-ref-15)
16. Payal Arora, ‘The Bottom of the Data Pyramid: Big Data and the Global South’, *International Journal of Communication*, 10 (2016): 1681–1699. [↑](#footnote-ref-16)
17. See: Roger Brownsword, Eloise Scotford, and Karen Yeung (eds), *The Oxford Handbook of Law Regulation and Technology*, Oxford: Oxford University Press, 2017. [↑](#footnote-ref-17)
18. See: Bert-Jaap Koops, ‘The trouble with European data protection law’, *International Data Privacy Law*, 4.4 (2014): 250-261. [↑](#footnote-ref-18)
19. See: Bert-Jaap Koops and Ronald Leenes, ‘Privacy regulation cannot be hardcoded. A critical comment on the ‘privacy by design’ provisions in data-protection law’, *International Review of Law, Computers and Technology*, 28.2 (2014): 159-171. [↑](#footnote-ref-19)
20. See: Tobias Matzner, ‘Why privacy is not enough privacy in the context of “ubiquitous computing” and “big data”’, *Journal of Information, Communication and Ethics in Society* 12.2 (2014): 93–106. [↑](#footnote-ref-20)
21. See: Mike Montiero, ‘Ethics can’t be a Side Hustle’, *Dear Design Student*, 19 March 2017,https://deardesignstudent.com/ethics-cant-be-a-side-hustle-b9e78c090aee ;Hamza Shaban, ‘Amazon Employees Demand Company Cut Ties with ICE’, *Washington Post*, 22 June 2018, https://www.washingtonpost.com/news/the-switch/wp/2018/06/22/amazon-employees-demand-company-cut-ties-with-ice/?utm\_term=.713db9c3092a; Sara Salinas and Jillian D’Onfro, ‘Google Employees: We no Longer believe the company places values over profit’, *CNBC,* 27 November 2018, https://www.cnbc.com/2018/11/27/read-google-employees-open-letter-protesting-project-dragonfly.html [↑](#footnote-ref-21)