



**Selected Papers of #AoIR2019:
The 20th Annual Conference of the
Association of Internet Researchers**
Brisbane, Australia / 2-5 October 2019

PRINCIPLES OF GOOD DATA

S.Kate Devitt
Defence Science & Technology Group

Monique Mann
Deakin University

Angela Daly
Strathclyde University

Introduction

In recent years, there has been an exponential increase in the collection, aggregation and automated analysis of information by government and private actors that disproportionately disadvantages the underrepresented, marginalized and unheard. In response to this there has been significant critique regarding what could be termed 'bad' data practices in the globalised digital economy. Examples include the Australian Centrelink 'robo-debt' action by the Australian government that affected thousands of vulnerable Australians (Medhora, 2019, Feb 28) or ongoing investigations into Facebook's use of private consumer data (Shechner & Secada, 2019, Feb 22). Considerations of 'bad data' practices often only provide *critiques* rather than engaging constructively with a new vision of how digital technologies and data can be used productively and justly to promote social, economic, cultural and politically progressive goals. In this paper we consider the fundamentals of Good Data to increase trust. We begin by conceptual considerations of the nature of 'data' and 'goodness' before defending fifteen principles of good data under four banners: *Community, Rights, Usability & Politics* in order to ultimately progress a more just digital economy and society.

What is data?

Grasping the nature of data and information assists understanding how information systems have been built, as well as how they ought to be designed and managed.

Suggested Citation (APA): Devitt, S., Mann, M., Daly, A. (2019, October 2-5). *Principles of Good Data*. Paper presented at AoIR 2019: The 20th Annual Conference of the Association of Internet Researchers. Brisbane, Australia: AoIR. Retrieved from <http://spir.aoir.org>.

Amongst a plethora of definitions of data and information, we align our principles with the Data Information Knowledge Wisdom (DIKW) model (Ackoff, 1989). DIKW puts 'data' at the bottom of an epistemological hierarchy beneath information, knowledge and wisdom (Zins, 2007). We accept the premise that all levels of information depend on an active agent. That is to say, information without an agent to *believe something* or to *do something* is either not information at all, or if it remains information, serves no purpose or value. Building on the DIKW model 'Data' can be interpreted as a component of an overall epistemic spectrum of information, a lower, more raw, unprocessed epistemic level of information, or a term being used as a proxy for the whole DIKW model. When evaluating data's 'goodness', it is worth considering which of these three uses is applicable. When we use the term 'data' in this paper we mean it as a term being used as a proxy for the whole DIKW model.

What is good?

How can data can be 'good'? Goodness can be a property of a thing, a service, a method, an event, a system, a process, a judgment, a sensation, a feeling or a combination of these. To identify 'good', we could suppose that there are moral facts (see Parfit 2011; Scanlon, 2014). Agreement on moral facts enables standards, policies, practices and frameworks to improve information systems and communicate expectations. However, given the limits of our knowledge of moral facts (should they exist) and in light of colonial and post-colonial data practices (Arora, 2016; Connell, 2007) we assume a hybrid moral theory—where we allow that some moral facts may be objective (e.g. 'tolerance' or 'openness') and others relative (e.g. Wong, 1984). A hybrid theory allows respect for cultural diversity and demands case-by-case determinations of goodness. By promoting a hybrid account, we are prepared for disagreement about what is good and assume that the discovery of moral facts (if they do exist) is non-trivial and unresolved. We advocate an ethic of active seeking, openness and tolerance to diverse views on 'the good' particularly consultation with the underrepresented, marginalised and unheard.

Principles of Good Data

Good data, particularly knowledge and wisdom in the DIKW model, contribute to understanding and justify progressive political action by collectives. 'Good data' is thus situated from an ethical perspective to *progress society*, rather than simply satisfying an epistemic goal to inform. Therefore, we connect Good Data with political action and social justice.

Building on our analysis of 'good data' we present 15 principles organised under four pillars: Community, Rights, Useability, Politics:

Community: Good Data must be orchestrated and mediated by and for data subjects (Principle 1), including communal sharing for community decision-making and self-governance (Principles 2 and 3).

Rights: Good Data should be collected with respect to humans and their rights and the natural world (Principle 4).

Usability: Good Data is usable and fit for purpose (Principle 5); consensual, fair and transparent (Principles 6, 7 and 8), and must respect interpersonal relationships (Principle 9). Good data is dependent on context, and with reasonable exceptions, should be open and published, revisable and form useful social capital (Principle 10).

Politics: Good data reveals and challenges the existing political and economic order (Principle 11) so that data empowered citizens can secure a good democracy (Principles 12, 13, 14, and 15).

Conclusion

Our research into 'good data' encourages data optimism beyond minimal ethical checklists and duties—thus our aim is supererogatory. We recommend good data principles to progress political and social justice agendas such as citizen-led data initiatives, accepting 'good enough' data to achieve aims. The aim is to dismantle existing power structures through the empowerment of communities and citizens via data and digital technologies. Moving away from the body of critique of pervasive 'bad data' practices by both governments and private actors in the globalised digital economy, we paint an alternative, more optimistic but still pragmatic picture of the datified future.

References

- Ackoff, R. L. (1989). From data to wisdom. *Journal of applied systems analysis*, 16(1), 3-9.
- Arora, P. (2016). The Bottom of the Data Pyramid: Big Data and the Global South. *International Journal of Communication* 10, 1681-1699.
- Cadwalladr, C. (2018, May 8). UK regulator orders Cambridge Analytica to release data on US voter. *The Guardian*. Retrieved from <https://www.theguardian.com/uk-news/2018/may/05/cambridge-analytica-uk-regulator-release-data-us-voter-david-carroll>
- Connell, R. (2007). *Southern Theory: The Global Dynamics of Knowledge in Social Science*. Crows Nest, Australia: Allen & Unwin
- Medhora, S. (2019, February 18). Over 2000 people died after receiving Centrelink robot-debt notice, figures reveal. *Triple J Hack*. Retrieved from <https://www.abc.net.au/triplej/programs/hack/2030-people-have-died-after-receiving-centrelink-robodebt-notice/10821272>
- Shechner, S. & Secada, M. (2019, February 22). You give apps sensitive personal information. Then they tell Facebook. *The Wall Street Journal*. Retrieved from <https://www.wsj.com/articles/you-give-apps-sensitive-personal-information-then-they-tell-facebook-11550851636>
- Parfit, D. (2011). *On What Matters*, Volume 2. Oxford: Oxford University Press
- Scanlon, T.M. (2014). *Being Realistic about Reasons*. Oxford: Oxford University Press.
- Wong, D.B. (1984). *Moral Relativity*. Berkeley, CA: University of California Press

Zins, C. (2007). Conceptual Approaches for Defining Data, Information, and Knowledge. *Journal of the American Society for Information Science and Technology*. 58(4), 479-93.