Data & Policy (2019), 1: e1, 1–5
doi:10.1017/dap.2019.2

EDITORIAL

Data & Policy: A new venue to study and explore policy–data interaction

Stefaan G. Verhulst[1], Zeynep Engin[2]* and Jon Crowcroft[3]

[1]The Governance Lab, New York University, New York, USA
[2]Department of Computer Science, University College London, London, United Kingdom
[3]Computer Laboratory, University of Cambridge, Cambridge, United Kingdom
*Corresponding author. Email: z.engin@ucl.ac.uk

Every era faces a unique set of challenges and dilemmas, but ours can credibly lay claim to some of the most complex and vexing that humankind may have ever confronted. From climate change to growing inequality to a rising tide of refugees: we face an intricate mesh of overlapping and interdependent difficulties, one that is pushing the limits of our existing policy and governance capabilities (Data for Policy, 2015; Meyer et al., 2017). What we require today are not so much (or not only) new solutions, but *new ways for arriving at solutions* (Susha et al., 2017). We need a twenty-first century paradigm of governance and policy making.

Data, it is increasingly clear, will be central to this paradigm (Pentland, 2013; Kirkpatrick, 2012). Along with ever increasing computer storage and analytics capabilities, massive amounts of data generated from citizens, devices, and sensors provide decision makers the opportunity to monitor and manage public infrastructure in real time and predict future patterns when used responsibly (Engin and Treleaven, 2019; Janssen and Helbig, 2018). Data have the potential to transform every part of the policy-making life cycle—agenda setting and needs identification; the search for solutions; prototyping and implementation of solutions; enforcement; and evaluation (Janssen and Helbig, 2018). These are all critical, interlinked steps in addressing our societal challenges, and each of these needs a radical rethink.

The idea that data could be a key differentiator is, of course, not a new one. Its potential has been evident for some time now (Wang et al., 2018), especially in the business world (Henke et al., 2016), but also in the policy community, where efforts to harness the power of information have yielded positive results in areas as disparate as gender equality (Fatehkia et al., 2018), improving urban traffic flows (Zhao et al., 2018), and enhancing regulatory compliance (Heat Seek, n.d.; Credit Suisse, n.d.). Successful data initiatives have been deployed by governments around the world in both developing and developed countries (Verhulst and Young, 2017a). Such initiatives have led to a growing recognition that data are and should increasingly be part of any effective governance toolkit.

Despite such encouraging results, it is true that the policy world has generally lagged behind business in its use of data and data methods (Hou et al., 2011). Policy–data interactions or governance initiatives that use data have been the exception rather than the norm, isolated prototypes and trials rather than an indication of real, systemic change. There are various reasons for the generally slow uptake of data in policymaking, and several factors will have to change if the situation is to improve. In particular, advocates of more data (and we include ourselves among this number) will need to overcome the following obstacles and limitations:

- Despite the number of successful prototypes and small-scale initiatives, policy makers' *understanding of data's potential and its value proposition generally remains limited* (Lutes, 2015). There is also limited appreciation of the advances data science has made the last few years. This is

© The Author(s) 2019. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

https://doi.org/10.1017/dap.2019.2 Published online by Cambridge University Press

a major limiting factor; we cannot expect policy makers to use data if they do not recognize what data and data science can do.

- The recent (and justifiable) backlash against how certain private companies handle consumer data has had something of a reverse halo effect: There is *a growing lack of trust* in the way data is collected, analyzed, and used, and this often leads to a certain reluctance (or simply risk-aversion) on the part of officials and others (Engin, 2018).
- Despite several high-profile open data projects around the world, much (probably the majority) of data that could be helpful in governance remains either *privately held or otherwise hidden in silos* (Verhulst and Young, 2017b). There remains a shortage not only of data but, more specifically, of high-quality and relevant data.
- With few exceptions, the *technical capacities of officials remain limited*, and this has obviously negative ramifications for the potential use of data in governance (Giest, 2017).
- It's not just a question of limited technical capacities. There is often *a vast conceptual and values gap* between the policy and technical communities (Thompson et al., 2015; Uzochukwu et al., 2016); sometimes it seems as if they speak different languages. Compounding this difference in world views is the fact that the two communities rarely interact.
- Yet, data about the use and evidence of the impact of data remain sparse. The impetus to use more data in policy making is stymied by *limited scholarship and a weak evidential basis* to show that data can be helpful and how. Without such evidence, data advocates are limited in their ability to make the case for more data initiatives in governance.
- Data are not only changing the way policy is developed, but they have also reopened the debate around *theory- versus data-driven methods in generating scientific knowledge* (Lee, 1973; Kitchin, 2014; Chivers, 2018; Dreyfuss, 2017) and thus directly questioning the evidence base to utilization and implementation of data within policy making. A number of associated challenges are being discussed, such as: (i) traceability and reproducibility of research outcomes (due to "black box processing"); (ii) the use of correlation instead of causation as the basis of analysis, biases and uncertainties present in large historical datasets that cause replication and, in some cases, amplification of human cognitive biases and imperfections; and (iii) the incorporation of existing human knowledge and domain expertise into the scientific knowledge generation processes—among many other topics (Castelvecchi, 2016; Miller and Goodchild, 2015; Obermeyer and Emanuel, 2016; Provost and Fawcett, 2013).
- Finally, we believe that there should be a sound under-pinning a new theory of what we call *Policy–Data Interactions*. To date, in reaction to the proliferation of data in the commercial world, theories of data management,[1] privacy,[2] and fairness[3] have emerged. From the Human–Computer Interaction world, a manifesto of principles of Human–Data Interaction (Mortier et al., 2014) has found traction, which intends reducing the asymmetry of power present in current design considerations of systems of data about people. However, we need a consistent, symmetric approach to consideration of systems of policy and data, how they interact with one another.

All these challenges are real, and they are sticky. We are under no illusions that they will be overcome easily or quickly.

They were the impetus behind the formation of the international Data for Policy conferences (dataforpolicy.org), launched in 2015. We were interested in initiating an interdisciplinary and cross-sector debate to bridge the gap between large-scale data-processing technologies and existing

[1] Data management and use: Governance in the 21st century—A British Academy and Royal Society project. https://royalsociety.org/topics-policy/projects/data-governance/.

[2] Privacy Enhancing Technologies. https://royalsociety.org/topics-policy/projects/privacy-enhancing-technologies/.

[3] Engineering a fair future: Why we need to train unbiased AI. Speaker: Dr Krishna Gummadi, 18, 20 and 21 February 2019, London, Manchester and Belfast. https://www.bcs.org/events-home/turing-talk/2019-turing-talk/.

https://doi.org/10.1017/dap.2019.2 Published online by Cambridge University Press

expert knowledge in major policy domains to make policy development processes more citizen-focused, taking into account public needs and preferences supported with actual experiences of public services. Since then we engaged in several parallel debates on the ethical and privacy concerns associated with such developments and the usability of technologies addressing the needs of diverse stakeholders.

During the past four conferences, we have hosted an incredibly diverse range of dialogues and examinations by key global thought leaders, opinion leaders, practitioners, and the scientific community (Data for Policy, 2015, 2016, 2017, 2019). What became increasingly obvious was the need for a dedicated venue to deepen and sustain the conversations and deliberations beyond the limitations of an annual conference. This leads us to today and the launch of *Data & Policy,* which aims to confront and mitigate the barriers to greater use of data in policy making and governance.

Data & Policy is a venue for peer-reviewed research and discussion about the potential for and impact of data science on policy. Our aim is to provide a nuanced and multistranded assessment of the potential and challenges involved in using data for policy and to bridge the "two cultures" of science and humanism—as CP Snow famously described in his lecture on "Two Cultures and the Scientific Revolution" (Snow, 1959). By doing so, we also seek to bridge the two other dichotomies that limit an examination of datafication and is interaction with policy from various angles: the divide between practice and scholarship; and between private and public.

Importantly, our intention is not simply to advocate for greater—and blind—use of data; while we recognize the very real possibilities, we also know that there are risks, and we believe that the ultimate goal is not simply data for the sake of data, but to arrive at a better understanding of how data can be used in an efficient and responsible manner to confront the challenges of our era. Therefore, while our pages will no doubt contain a fair number of authors who advocate the use of data in governance, readers can also expect more nuanced and even skeptical perspectives.

We also see the potential with *Data & Policy* to extend beyond the idea of a conventional academic journal. The movements towards more open, transparent, and collaborative research—including the sharing of materials not typically published in academic journals—are highly relevant to our project of linking technical, policy, and other expertise and for building trust. Articles published in *Data & Policy* will be open access: freely available under licensing that allows unimpeded reading, sharing, and reuse, helping us to reach readers and potential authors in academic institutions, government agencies, international, nonprofit, and commercial organizations, and the general public. Beyond this we will encourage the open availability of data, code, and other materials to promote transparency and reuse, albeit recognizing that there are circumstances where this is not possible or responsible. Authors submitting to *Data & Policy* will be asked to provide a data availability statement that either links to the data underlying the results and other relevant materials or that explains the reasons why these cannot be shared. We are conscious of the need to support authors in this process so we provide information about the different resources that can be used. We encourage authors to think beyond traditional outputs to also share proposals, posters, presentations, and policy-related problems that require investigation.

To try and address the terminology and conceptual gaps that exist between different communities, we will also seek to innovate with the formats published in *Data & Policy* and the features within them. Articles will be published with a short but prominent policy significance statement to summarize their relevance to policy makers in language that is understandable to the wider public. We are actively soliciting ideas from the Data for Policy and the *Data & Policy* audiences about the types of content that can help us bridge the communities we are appealing to.

It is essential to say one more thing. *Data & Policy* is about policy making; it is not about politics. Throughout our enquiries, we will strive to remain ideologically neutral and avoid the political schisms that define so much of public life and discourse nowadays. This does not mean that we are unaware of the social and political contexts within which our papers are written (and will be received), but it does mean that our aspiration is to remain pragmatic and results-oriented. We seek to discover what works and how to replicate successful data initiatives at a larger scale or in different geographies.

So these are our principles: scholarly, pragmatic, open-minded, interdisciplinary, focused on actionable intelligence, and, most of all, innovative in how we will share insight and pushing at the

https://doi.org/10.1017/dap.2019.2 Published online by Cambridge University Press

boundaries of what we already know and what already exists. We are excited to launch *Data & Policy* with the support of Cambridge University Press and University College London, and we're looking for partners to help us build it as a resource for the community. If you're reading this manifesto it means you have at least a passing interest in the subject; we hope you will be part of the conversation.

Join us by reading and publishing in *Data & Policy* (cambridge.org/dap), and follow us on social media (@data_and_policy).

References

Castelvecchi D (2016) Can we open the black box of AI?. *Nature News* published online 5 October 2016, https://www.nature.com/news/can-we-open-the-black-box-of-ai-1.20731.

Chivers T (2018) How big data is changing science." *Mosaic* published online 2 October 2018, https://mosaicscience.com/story/how-big-data-changing-science-algorithms-research-genomics/.

Credit Suisse (n.d.) *How Big Data Analytics Is Transforming Regulatory Compliance*. Available at https://www.credit-suisse.com/corporate/en/articles/news-and-expertise/how-big-data-analytics-is-transforming-regulatory-compliance-201711.html (accessed 9 May 2019).

Data for Policy (2015) Policy-making in the Big Data Era: Opportunities and Challenges, University of Cambridge, 15–17 June 2015. Available at http://dataforpolicy.org/data-for-policy-2015/.

Data for Policy (2016) Frontiers of Data Science for Government: Ideas, Practices, and Projections, University of Cambridge, 15–16 September 2016. Available at http://dataforpolicy.org/data-for-policy-2016-2/.

Data for Policy (2017) Government by Algorithm? London–Westminster Conference Centre (1VS), 6–7 September 2017. Available at http://dataforpolicy.org/data-for-policy-2017/.

Data for Policy (2019) Digital Trust and Personal Data, University College London, 11–12 June 2019.

Dreyfuss E (2017) Want to make it as a biologist? Better learn to code. *Wired* published online 10 March 2017, https://www.wired.com/2017/03/biologists-teaching-code-survive/.

Engin Z (2018) Digital ethics: data, algorithms, interactions. *Zenodo* published online, https://doi.org/10.5281/zenodo.1689674.

Engin Z and Treleaven P (2019) Algorithmic government: automating public services and supporting civil servants in using data science technologies", *The Computer Journal* **62**(3), 448–460. https://doi.org/10.1093/comjnl/bxy082.

Fatehkia M, **Kashyap R and Weber I** (2018) Using Facebook ad data to track the global digital gender gap. *World Development* **107**(July), 189–209. https://doi.org/10.1016/j.worlddev.2018.03.007.

Giest S (2017) Big data for policymaking: fad or fasttrack? *Policy Sciences* **50**(3), 367–382. https://doi.org/10.1007/s11077-017-9293-1.

Gil-Garcia JR, **Chun SA and Janssen M** (2009) Government information sharing and integration: combining the social and the technical. *Information Polity* **14**(1,2), 1–10. https://doi.org/10.3233/IP-2009-0176.

Heat Seek (n.d.) Heat Seek. Available at https://heatseek.org/ (accessed 9 May 2019).

Henke N, **Bughin J**, **Chui M**, **Manyika J**, **Saleh T**, **Wiseman B and Sethupathy G** (2016). *The Age of Analytics: Competing in a Data-Driven World*. McKinsey. Available at https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-age-of-analytics-competing-in-a-data-driven-world.

Hou Y, **Lunsford RS**, **Sides KC and Jones KA** (2011) State performance-based budgeting in boom and bust years: an analytical framework and survey of the states. *Public Administration Review* **71**(3), 370–88. https://doi.org/10.1111/j.1540-6210.2011.02357.x.

Janssen M and Helbig N (2018) Innovating and changing the policy-cycle: policy-makers be prepared!. *Government Information Quarterly* **35**(4, Supplement), S99–S105. https://doi.org/10.1016/j.giq.2015.11.009.

Kirkpatrick R (2012) Big data for development. *Big Data* **1**(1), 3–4. https://doi.org/10.1089/big.2012.1502.

Kitchin R (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, https://doi.org/10.1177/2053951714528481.

Lee DB (1973) Requiem for large-scale models. *Journal of the American Institute of Planners* **39**(3), 163–178. https://doi.org/10.1080/01944367308977851.

Lutes T (2015) *Data-Driven Government: Challenges and a Path Forward*. IBM. Available at https://www.ibm.com/downloads/cas/BJE8Z48J?mhq=data%20driven%20government&mhsrc=ibmsearch_a.

Meyer ET, **Crowcroft J**, **Engin Z and Alexander A** (2017) Data for public policy. *Policy & Internet*, https://doi.org/10.1002/poi3.147.

Miller HJ and Goodchild MF (2015) Data-driven geography. *GeoJournal* **80**(4), 449–461. https://doi.org/10.1007/s10708-014-9602-6.

Mortier R, **Haddadi H**, **Henderson T**, **McAuley D and Crowcroft J** (2014) Human-Data Interaction: The Human Face of the Data-Driven Society. Available at https://ssrn.com/abstract=2508051 or http://dx.doi.org/10.2139/ssrn.2508051.

https://doi.org/10.1017/dap.2019.2 Published online by Cambridge University Press

Obermeyer Z and Emanuel EJ (2016) Predicting the future—big data, machine learning, and clinical medicine. *The New England Journal of Medicine* **375**(13), 1216–1219. https://doi.org/10.1056/NEJMp1606181.

Pentland A (2013) The data-driven society. *Scientific American* **309**(4), 78–83.

Provost F and Fawcett T (2013) Data science and its relationship to big data and data-driven decision making. *Big Data* **1**(1), https://doi.org/10.1089/big.2013.1508.

Snow CP (1959) *The Two Cultures and the Scientific Revolution*, New York: Cambridge University Press.

Susha I, **Janssen M and Verhulst S** (2017) Data collaboratives as a new frontier of cross-sector partnerships in the age of open data: Taxonomy development. *50th Hawaii International Conference on System Sciences*, https://doi.org/10.24251/HICSS.2017.325.

Thompson K, **Daly C**, **Keene C**, **Raj M and Symons R** (2015) Bridging the gap between data and policy: Engaging national governments in youth development. *Deloitte Insights*, Available at https://www2.deloitte.com/insights/us/en/industry/public-sector/engaging-governments-in-youth-development.html.

Uzochukwu B, **Onwujekwe O**, **Mbachu C**, **Okwuosa C**, **Etiaba E**, **Nyström ME and Gilson L** (2016) The challenge of bridging the gap between researchers and policy makers: Experiences of a Health Policy Research Group in engaging policy makers to support evidence informed policy making in Nigeria. *Globalization and Health* **12**(November), https://doi.org/10.1186/s12992-016-0209-1.

Verhulst SG and Young A (2017a) *Open Data in Developing Economies: Toward Building an Evidence Base on What Works and How.* African Minds. Available at http://www.africanminds.co.za/dd-product/open-data-in-developing-economies-toward-building-an-evidence-base-on-what-works-and-how/.

Verhulst SG and Young A (2017b) *The Potential of Social Media Intelligence to Improve People's Lives*. The Governance Lab. Available at http://www.thegovlab.org/static/files/publications/social-media-data.pdf.

Wang Y, **Kung LA and Byrd TA** (2018) Big data analytics: understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change* **126**(January), 3–13. https://doi.org/10.1016/j.techfore.2015.12.019.

Zhao Y, **Zhang H**, **An L and Liu Q** (2018) Improving the approaches of traffic demand forecasting in the big data era. *Cities* **82** (December), 19–26. https://doi.org/10.1016/j.cities.2018.04.015.

Cite this article: Verhulst, S. G. Engin, Z. and Crowcroft, J. 2019. *Data & Policy*: A new venue to study and explore policy–data interaction. *Data & Policy* 1: e1, 1–5. https://doi.org/10.1017/dap.2019.2

https://doi.org/10.1017/dap.2019.2 Published online by Cambridge University Press