

Making the Invisible Visible: Enacting a Vision for Augmented Intelligence

Theresa Dirndorfer Anderson
University of Technology Sydney (UTS)
Sydney, Australia
Theresa.Anderson@uts.edu.au

Rachael Fraher
NSW Data Analytic Centre
Sydney, Australia
Rachael.Fraher@treasury.nsw.gov.au

Data-intensive environments compel us to address the challenges and complexities raised in studies of cyberinfrastructure (Bowker et al., 2010), including ethical considerations about the capacity for individuals to resist machine-generated classifications and characterizations or offer alternative interpretations (boyd & Crawford, 2012; Ratto, 2007). Concerns that data and AI-informed technologies truly serve humans (as individuals and collective groups) is fuelling government and social consideration of systems and functions that may need to be created to mitigate the damage caused by information and data asymmetries. Within the wider community, there is a growing consciousness about the vulnerability of data to misinterpretation, misuse and misappropriation. In this context, raising questions about how social values can be 'baked' into data infrastructure grows in importance.

This paper explores ways social informatics is shaping the data practices of a government data agency. It reports on the theory/practice collaboration of the two authors, reflecting on progress made shaping a strategy for gaining public trust in the agency's data practices. This paper shares how embedding a social informatics framework into training designed for these analytics professionals is sensitising them to what Meyer (2014) refers to as the significance of the "hyphen" – the co-evolving relationship between people and data technologies -- for their work practices.

Building bridges to support transdisciplinary data practices

The data agency at the heart of this project was established in response to growing government interest in the potential of data analytics contributing to better outcomes for communities by generating new perspectives on complex problems. The agency's primary mission involves building world-class capabilities in whole-of-government data analytics to support the government's efforts to improve health and well-being, safety, social, economic and environmental outcomes for its citizens. As a centralised agency it has the capacity to aggregate and analyse whole of government data, and serves as a single port of call for departments seeking to leverage data sets for their own purposes. The multidisciplinary team collaborates with other government agencies on complex, stakeholder-oriented problems in a broad range of domains: road safety, small enterprise innovation, urban infrastructure, environment, public health, and child welfare to name a few. Particularly in this last example, the data science work performed often involves complex and highly sensitive data sets. The agency has become a trusted data user and a source of technical, methodological and practice expertise serving as a model for other jurisdictions.

Our ongoing collaboration began with an invitation from the agency director to develop co-design frameworks and participatory models for working with vulnerable populations within their jurisdiction. Very quickly it emerged that the workplace challenge for this technologically talented team was similar to the situations described by Wallach (2014), who points to the insight that social scientists can offer to machine learning projects, and Kitchin (2014), who points out that "new data analytics" needs not only new forms of data management and analytical techniques but also boundary-spanning epistemological reframings. For this project, we frame this challenge as an ethical data practice of learning to 'make the invisible visible' by remaining alert to who (and what) is

missing, under-represented or mis-represented in the data. To nurture a more nuanced view of data that sensitises to the limitations of assumptions that the data available represents reality, we sought to build disciplinary bridges for data science practitioners in the agency.

Moving from data-driven to data-informed

When pressing for greater fairness, accountability and transparency in data science practices, Wallach (2014) points out that big data is essentially granular social data. There is thus an imperative for data modeling frameworks to retain a human touch. Framing complex data infrastructure as a socio-technical system explodes assumptions of data-driven “actionable insights” emerging from a single source of (data) truth and sets up an opportunity to demonstrate the value-add of this human touch in the lifecycle of any data project.

To raise practitioner awareness of the potential limitations of ‘data-driven’ decision making, every opportunity is sought to deliberately disrupt assumptions about data objectivity and human subjectivity. To do so this project draws on Anderson’s work on creative information ecologies (2013) and approach to uncertainty as a delightfully human construct (2010). Intuition and emotion are as inevitable a part of that decision making process as the uncertainty inherent in any data, no matter the scale.

Marcia Bates insight within her 1990 paper asking “Where should the person stop and the information search interface start?” was particularly compelling for promoting an augmentation approach to the practitioners’ computational work.

If we use, rather than ignore, the special traits of humans in the design of human-computer interfaces for information systems, we may find our abilities enhanced in unpredictable and creative ways (p 590)

It is a position reminiscent of Doug Engelbart’s (1962) vision for computational design augmenting human intelligence suggests

...a way of life in an integrated domain where hunches, cut-and-try, intangibles, and the human “feel for a situation” usefully co-exist with powerful concepts, streamlined terminology and notation, sophisticated methods, and high-powered electronic aids.

Interpretation drawing on intuition and the insight born of lived experience is something humans do best.

The decisions we make as a consequence of data input are based on our value judgments about that data – judgments that are subjective and emotional rather than rational. This need not be problematic. For data analytics professionals more accustomed to framing these human factors as obstacles rather than enablers of their computational work, however, there is a great deal of “unlearning” involved. When engaging with these data professionals, we are therefore always looking for ways to change their languaging of their practice from “data driven’ to ‘data-informed’.

Making the invisible visible

In our project, this unlearning takes place by making visible the value-laden chain of practices and policies shaping the insights gleaned from data. The project draws particularly on sociotechnical explorations of classification, categorisation and infrastructure studies (e.g. Bowker and Star, 2000; Bowker et al.,2010; Edwards et al, 2009; Star & Ruhleder, 1996). The inherent power of the politics of classification and information structures to magnify inequities in technology advances can be countered if we train people all along the data lifecycle to remain alert for the missing, misrepresented and under-represented in our data and information structures.

Social informatics provides a framework for foregrounding critical data ethics considerations and engagement with the background “shadow work” (Sawyer & Tapia, 2006; Star & Strauss, 1999) of data practices within the agency’s organisational and political contexts. To embed awareness of the consequences of choosing specific technologies for their practice (Kling, 2007), the training designed as part of this project intentionally draws attention to the infrastructures within which data practices unfold. Accepting the social actor approach to understanding our engagements with systems articulated in Lamb & Kling (2003) means that the social complexities of working together and in organizational settings must be accounted for even as technical complexity increases. Recognising that data practices are embedded in a larger ecology of social relations, physical arrangements, cognitive engagements, and technological elements, according to Jarrahi et al. (2019), focuses analytic attention to the connections. Critical reflection on these socio-technical entanglements sheds light on the relations of power and invisibility and the implications of varying perceptions of visibility for their practice. Engaging with the sociotechnical complexity of such data assemblages makes us acutely aware of how substrate can become substance (Star and Ruhleder, 1996), drawing attention to the very human features of working with data and information in the “web of computing” (Kling & Saatchi, 1982).

Providing a socially rich perspective of data assemblages

A core message when engaging with these data professionals is that data does not speak for itself, but rather, is given a voice by the people and the algorithms that play such critical roles in the transformation of data into insight. If unchecked, the assumptions and values embedded in the algorithms and the decisions they drive sink into everyday information infrastructure. The assemblages that evolve in this human-machine-information interplay rarely lend themselves to deliberate design and yet, ironically, as they become more naturalized and more invisible, their configuration can become more frozen.

The paper presents a working model devised as a consequence of this collaboration with the data agency that provides a sociotechnical framework for ethical data practice framed around four key interwoven components: data, design, people, policy. In effect this framework is serving as an aid for the “discovery processes” (Kling, 2007) that will contribute to the integration of participatory models and mechanisms for ongoing feedback with the community with data management and data sharing regimes in place within the agency. Tools are made through practice (Engeström, 1990). In helping these practitioners to appreciate ways that human judgments and values can and should be permitted to augment the computational components of the data assemblages they are shaping, the human capacity to work with uncertainty and intuitive judgements comes to be seen as an essential partner to high-powered computational and analytical capacities.

Engaging with the “data publics” represented in the data sets in use within this government agency is also critical for evaluating the trust-building work undertaken to date. To this end, the next step of our project will involve engaging more directly with the *people* and *policy* components of the framework devised for this project to examine if the social informatics analysis and “socially rich” understandings of these data assemblages will have repercussions for the everyday life of the citizens in this agency’s jurisdiction.

References

- Anderson, T.D. (2013). The 4Ps of innovation culture: conceptions of creatively engaging with information. *Information Research*, 18(3) paper C28. Available at <http://InformationR.net/ir/18-3/colis/paperC28.html>.
- Anderson, T.D. (2010) Kickstarting creativity: supporting the productive faces of uncertainty in information practice, *Information Research*, 15(4), paper colis721. Available <http://InformationR.net/ir/15-4/colis721.html>.
- boyd, d., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662-679.
- Bowker, G. C., Baker, K., Millerand, F., & Ribes, D. (2010). Toward information infrastructure studies: Ways of knowing in a networked environment. In J. Hunsinger, L. Klastrup, & M. Allen (Eds.), *International handbook of internet research* (pp. 97-117). Dordrecht, Netherlands: Springer.
- Bowker, G. C., & Star, S.L. (2000). *Sorting things out: Classification and its consequences*. MIT Press.
- Edwards, P. N., Bowker, G. C., Jackson, S. J., & Williams, R. (2009). Introduction: an agenda for infrastructure studies. *Journal of the Association for Information Systems*, 10(5), 364-374.
- Engelbart, D.C. (1962) Augmenting human intellect: a conceptual framework . SRI Summary Report AFOSR0 3223, Available <http://www.dougenelbart.org/pubs/augment03906.html>.
- Engeström, Y. (1990) When is a tool? Multiple meanings of artifacts in human activity. In *Learning, Working and Imagining: Twelve studies in activity theory* (pp 171-195), Helsinki, Finland: Orienta-Konsultit Oy.
- Jarrahi, M.L., Philips, G., Sutherland, W., Sawyer, S., Erickson, I. (2019) Personalization of Knowledge, Personal Knowledge Ecology, and Digital Nomadism, *Journal of the Association for Information Science and Technology* 70(4):313–324, 2019
- Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1).
- Kling, R. (2007). What Is Social Informatics and Why Does It Matter? *The Information Society*, 23(4), 205 – 220.
- Kling, R. & Saatchi, W. (1982) The web of computing: Computing technology as social organization, *Advances in Computers*, 21: 3078.
- Lamb, R. & Kling, R. (2003) Reconceptualizing users as social actors in information systems research. *MIS quarterly*, 27 (2), 1970236.
- Meyer, Eric T. (2014). Examining the Hyphen: The Value of Social Informatics for Research and Teaching. In Pnina Fichman and Howard Rosenbaum (Eds.), *Social Informatics: Past, Present and Future* (pp. 57-74). Cambridge, UK: Cambridge Scholarly Publishers.
- Ratto, M. (2007). Ethics of seamless infrastructures: Resources and future directions. *International Review of Information Ethics*, 8 (12), 20027.
- Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information systems research*, 7(1), 1110134.

Sawyer, S., & Tapia, A. (2006). Always articulating: Theorizing on mobile and wireless technologies. *The Information Society*, 22(5), 311-323.

Star, S. L., & Strauss, A. (1999). Layers of silence, arenas of voice: The ecology of visible and invisible work. *Computer supported cooperative work (CSCW)*, 8(1), 9-30.

Wallach, H. (2014) Big Data, Machine Learning, and the Social Sciences: Fairness, Accountability, and Transparency, *Medium*, <https://medium.com/@hannawallach/big-data-machine-learning-and-the-social-sciences-927a8e20460d>