Design Principles and Tensions for Collaborative Digital Governance in an Artificial Intelligence Age

Paul W. Fay Henman(✉)

School of Social Science, The University of Queensland, Brisbane, Australia
p.henman@uq.edu.au

Abstract. Complex policy problems seem to be increasing, such as pandemics, climate change, poverty and inequality, and financial and political instability. Such challenges require collaborative and intelligent policy and administrative responses. Innovations in digital technology, such as big data analytics and artificial intelligence (AI), are routinely advocated as providing solutions. However, technology itself cannot solve these complex problems and can make matters worse. Rather, thoughtful, and deliberative approaches to designing collaborative digital administrative systems are needed; ones that appreciate both the strengths and weaknesses of digital technology and recognize the ongoing importance of humans in socio-technical administrative practices. This plenary paper will begin by outlining the promise of collaborative digital governance and some of the key emerging technologies being used by governments. Then, by examining several successful and unsuccessful examples of digital government projects, the paper will articulate a series of lessons in the form of design principles and tensions to balance. It is hoped that these design principles will assist public administration leaders to build the administrative solutions that challenge governments to enable society to steer towards shared futures.

Keywords: Governance · Collaboration · Artificial Intelligence · Design

1 Introduction

We are currently living in a period of much change and uncertainty. It seems that the policy challenges that governments are being asked to manage grow more complex and urgent each year. In the last few years alone, we have witnessed a global pandemic (COVID-19), escalating climate change, growing numbers and intensity in disasters, transnational disruptions in supply chains, major economic disturbances, shifting balances and behavior among global superpowers, growing disinformation circulating through the internet and social media, and plummeting trust in public institutions. All these challenges place great expectations on states – their political and policy administration leaders – to have solutions, to manage the disruptions so that society and the economy continue to function to achieve collective ends.

A. Rahman et al. (Eds.): ICAS 2022, ASSEHR 776, pp. 406–416, 2024.
https://doi.org/10.2991/978-2-38476-104-3_39

From Rittel and Webber [1], such challenges are known as 'wicked problems', those intractable issues that bypass standardized rational, technical thinking. As Head [2] points out they are characterized by uncertainty, complexity, and value divergence among multiple actors, and where even the problem may not be well understood. Wicked problems necessarily need political management working in hand with policy analysts and developers, ideally by building partnerships, managing stakeholder and value conflicts, tackling knowledge uncertainties, and investing in preventative approaches. Collaborative governance is at its heart.

It is in this politico-administrative space that digital technology enters. Like the proverbial snake oil salesman offering simple solutions to complex problems, digital technology has long been touted as solving public policy and administrative challenges, or simply making things better. Digital technology seems to operate just like "magic" [3]. To be sure, digital technology can deliver benefits for enhancing administrative processes and particularly expanding what is possible beyond human only administration, but to achieve positive outcomes requires critical understanding of the strengths and limitations of digital technology. Indeed, digital technology cannot solve problems that are political and social in nature.

2 Digital Technology and Public Governance

For over 50 years digital (or electronic or information) technologies have progressively entered public administrations and increasingly expanded their involvement in diverse activities and roles [4].

Databases form the bedrock of contemporary administration in replacing paper files and filing cabinets to store administrative data. Human entry of data to these databases via keyboards is being supplemented with automated sensors (e.g., pollution monitors) and circulation of data from one database to another via a global network of interconnected data exchanges. Once in databases, digital technology is crucial for calculation, analyzing and automating, with growing automated decision making of administrative decisions. Digital technologies can provide support to human administrative decision makers through structuring administrative process flows, providing joined-up data, classifying information, and predicting outcomes. Computer modelling and digital twins (digital versions of real structures) help policy makers and administrators make sense of the world and to consider alternative scenarios. Natural language processing technologies enable chatbots, as artificial conversational administrative agents.

Reflecting the flexibility of digital technology, the sheer diversity of how and where digital technologies are being used and being envisaged for use is incredible. Perri [5] delineates four different domains: administration; service delivery; democratic processes; and governance.

Apart from the domains in which digital technology is deployed there is also variety in the way they are created in relation to human administrators. A dominant underlying objective of senior administrators is to automate human tasks – either with a view to replace or deskill workers. This is the view unpinning Bovens and Zouridis's argument that we are seeing a shift from street level bureaucracies to screen-level to system-level bureaucracies [6]. Alternatively, digital technology can be used to supplement, augment, and extend human operations [7].

The visions of how digital technologies are used in administration is intimately wrapped into visions of administration and administrative agencies, and their role in public governance. Often administrators make the mistake that digitizing administrative processes is simply administration by other means. Rather, digitizing administration re-constitutes administrative practices and can operationalize new administrative principles by default (whether intended or not), and also bringing in new institutional relationships with the state through technological transfer and contractual relationships with the computer industry.

3 Digital Administrative Transformations

To help illustrate some of the administrative opportunities and challenges of digital technology in administration, this section briefly summarizes three case studies.

3.1 Digital Identification: India's Aadhaar Card

In many developing countries, a key challenge in delivering public services is identification of people who might be eligible to receive the service and to confirm a person's identity. Indeed, many people are invisible to government systems leading to poor knowledge about a country's residents and their situations and circumstances. Digital technology is being used to create such visibility, to make people legible so governments can "see like a state" [8], through the creation of digital identity systems and identity cards. India's Aadhaar Card is an example of such a system.

In India, the Aadhaar identity card incorporates biometric (i.e., fingerprints) provides proof of residence but not citizenship. Established in 2009, it is regarded as constituted the largest biometric ID system in the world. It is used to help access to public services and can be used to with banks and SIM cards. It has been an important instrument to enable the Indian government to identify and deliver welfare benefits to those who are eligible [9]. In making visible India's residents it forms a big dataset that enables public governance and planning.

While there have been significant benefits from the use of the card, there has been considerable concerns about surveillance and control of citizens and data security challenges [10]. The process of enrolling and receiving an Aadhaar card has experienced challenges resulting in exclusion from accessing services. Central to the card is scanning and uploading a fingerprint signature. However, many poorest of the poor have been found that they cannot register because of worn off fingerprints or missing fingers, thus resulting in lack of identity [11]. This is despite the card being voluntary and legal cases finding that the card is not necessary to access services a resident is entitled to receive.

3.2 Automated Decision Making: Australia's Robodebt Context

Using digital technology to conduct calculations – such as taxation rate and benefit levels – and to assess eligibility for services or benefits based on clear cut data has been widely used for decades. These tools are typically based on well defined legal requirements. Thus, digital technology automated policy, and indeed code becomes policy. Such automation is not without problems, as Australia's Robodebt system illustrates.

Introduced in late 2016, Robodebt – or technically the Online Compliance Initiative – involved automating a previously hybrid human-automation process of identifying possible overpayments of social security benefits and recovering subsequent debt [12]. The system began by matching annual income data in the taxation system with fortnightly data in the social security system to identify potential discrepancies. If discrepancies were detected, then the system would request social security recipients to provide evidence of income. Subsequently automated debt notices were issued.

Following considerable public debate, Parliamentary inquiries, and legal cases, the courts found, and the Australian government accepted that income averaging of annual income was an unlawful basis for calculating and pursuing debt. Also criticised was a key administrative principle of the burden for the onus of proof was reversed, from the government to the citizen, compounded by Robodebt's lack of transparency in how alleged debts were created [13, 14].

3.3 Predictive Analytics: Allegheny County's Family Screening Tool

Digital technology is increasingly being used for predictions, risk scoring and associated classification. Such technology has traditionally been based on traditional regression statistical analyses, but machine learning has opened up the potential for greater differential and accuracy. Predictive tools are increasingly being used in policing and criminal justice, child protection, employment services and taxation.

The Family Screening tool used by Allegheny County, Pennsylvania since 2016 is one example of predictive analytics. It is used by call centre staff receiving a possible notification of child abuse/neglect. Based on an analysis of large number of datasets, the tool seeks to predict the long-term likelihood of future involvement in child welfare by classifying a case on a 20-point scale [15]. Analysis suggests and an ethical assessment [16] suggests that the screening tool makes better predictions than humans allowing the County to better focus resources to those children at most risk. Though there have also been criticisms [17] that the system may reproduce racial, and class based biased intervention, does not take account of positive changes in parental ability to care, can counter-productively induce parents to seek help when they need it, and lead to workers simply acting on the tool rather than exercising professional judgement. Further, as child protection is a high-risk setting, misallocation errors can have the potential to create poorer outcomes, and potentially drive greater criminalisation rather than service provision.

4 Design Principles for Collective Governance

In designing digital collective governance to address wicked problems, consideration needs to be given to the digital technology, the nature of digital data, and the organizational context. I outline several design principles in each of these three domains.

4.1 Technology

The first design principle is: **Digital technology creates new possibilities**. Digital technology does not simply automate human actions, but extends the realm of the possible,

the scope of state action. We need to think beyond the mantra of efficiency gains and greater administrative accuracy. My own work from 30 years ago found that computerising Australia's social security system did not result in reduced administrative expenditure or even staff, but enabled greater complexity of policy and administration, more differentiated approaches to service delivery, and enhanced compliance [18]. Upgrading a road to a highway both enhances traffic along the highway, it disrupts traffic flows across it. Similarly, a paradoxical corollary of design principles one is that introducing digital technology also closes off possibilities. Digital technology is an infrastructure that has its own momentum and friction. Indeed, a former UK Chancellor admitted that he was unable to increase COVID support as much as he had wanted due to the design limitations of its computer systems [19], which is somewhat in conflict with the ubiquitous image of digitisation creating greater flexibility and agility.

Following from the first design principle is that **digital technology is not neutral**. It is not a simple, inert administrative device. The algorithms and associated data of a digital device embeds particular modes of understanding and operation, social relations and politics [20, 21]. Sometimes these politics are overt and purposeful, as was evident in Australia's Robodebt scheme, but often they are subtle and unintended, enacted from the way in which designers and programmers imagine the world, the organisation, and administrative practices.

Our third design principle is to **think beyond intended and unintended consequences**. Digital technologies are designed with purposes, visions of how they are to function and in relation to humans. Too often digital technology acts in other ways, and people use the technology to different purposes, consequences occur, that the designers do not consider. Designers need to be more imaginative and consider what the materiality of their designs may bring about beyond their plans. As Virilio stated, "When you invent the ship, you invent the shipwreck" [22]. Designers must prepare for both the ship and the shipwreck.

Fourthly. **Digital technology is not *the* solution,** nor is it at the centre of the solution. This is particularly the case with wicked problems. Digital technologies can facilitate collaboration through data sharing, collective communication, helping collaborators to visualise the problem, and so forth. It can provide avenues to collect data to better understand the problem – be that from humans or automated sensors. Administrators need to begin by understanding the domain in which they are working and identifying possible solutions – some of which may involves digital technology and others will have no digital technology.

Following from the fourth principle, the fifth design principle is: **High tech does not equal best tech.** Stated another way, the newest, most sophisticated digital technology is not often the best technological solution. Rather, administrators need to identify the right tool for the job, which could be mundane and old fashioned, like paper and people. Globally there is both great excitement and great worry about Artificial Intelligence, which is usually meant to refer to Machine Learning algorithms [23]. Machine learning algorithms can provide enhanced capacity to process data and contribute to administrative decisions, however the administrative design considerations is more about the role technology is going to play in administration rather than about the technical sophistication. For example, automating legal decisions that are well defined by law is best served

by traditional hard coded algorithms, as in the case of Robodebt. Building digital infrastructure to manage a domain will typically rely on mundane databases. Meanwhile, using machine learning for predicting child abuse/neglect cases *may* be more accurate, but not change the fundamental administrative principles and considerations that administrators need to consider ensuring responsible and accountable governance.

The sixth design principle follows: **Do not start with the technology**. Often computer or technology companies will offer their tools as solutions for a purported organisational problem or suggest it will enhance organisational or administrative performance. Do not take this bait. Begin instead with identifying the problem, or the organisational or administrative vision you are trying to achieve, and then identify what the solution might be. If you start with the technology, you are like Heidegger's proverbial worker with a hammer. In *Being and Time* [24] Heidegger explained that if the only tool you have is a hammer, then the world will only look hammerable. Technology is not simply an inert, fully formed thing that just gets dropped into an organization setting. As Fountain [25] emphasises, digital technology is *enacted* in an organisational setting. Digital technology is configured to its organisational location and practices. The reverse is also true, organisational processes and practices are configured for the technology. Hence, designing and building digital technology for collaborative governance requires is a socio-administrative-organisational-technical exercise. We are not building digital technology but building administrative systems and processes which incorporate digital technology. Thus, technologies must mesh with organisational practices. Thus, administrative and governance transformation must start with strategy, not technology [26]. The failure to understand this underpins much digital government project failures and overruns, as the technology is a wrong fit, people do not use it, or organisations deploy workarounds.

4.2 Data

Data is the second domain of digital design principles for collaborative governance. Central to collaborative governance is the need for sharing data across many organisations and sectors in order to reach an understanding of the nature and dynamics of wicked problems. Data also provides the means for modelling and assessing experimental policy and administrative responses.

The seventh digital design principle is that **big data is beneficial only if it reliable**. Big data – having huge volume, being large variations in type, and often being generated at high speed – provide qualitatively new ways for understanding the world and its dynamics. Big data is required for machine learning, and thus artificial intelligence [27]. At the same time, just as the old computer science adage state 'garbage in, garbage out' (GIGO), big data is not useful if it is no reliable or has veracity [28].

Data reflects the world and its creation Digital data is an outcome of its social construction. Data has a social life. Similar to the algorithms on which digital technology is built, data reflects the world in which it has been generated. We all have digital selves, or digital doppelgangers, that are created from our data traces. Designers, however, need to be aware of their role in creating the digital data worlds the datasets will constitute. Data thus have specific data ontologies, ways of seeing the world, that constitute realities in

data [29]. Thus, the way an algorithm defines what data is collected and the format of that data – e.g., is gender defined as a dichotomous variable, trichotomous, or something else – constitutes our digital knowledge. Indeed, the digital nature of digital technology tends to shape digital data as well structured, discrete, often quantitative [30, 31]. As the example of India's Aadhaar Cards illustrates, designers also need to be aware of the real-world conditions in which digital data is generated. Perhaps the human or technological inputters are faulty, or that the variable 'failure to cooperate' does not mean what it says [17]. Indeed, we may create data about ourselves, particularly on social media platforms, that provide a spin or a gloss or a mirage of how we want others to see ourselves [32].

It follows then that the ninth design principle is that **data are not a full reflection of the world**. Our digital doubles are not us. They may accurately reflect some of our characteristics and what we have done, but they are not us. Administrators thus need to be alert to the ways in which data is incomplete picture of what we might need to know. It could be missing parts, misleading, or false. Digital data is therefore, not neutral, nor objective.

Our tenth design principle is **manage the tension between holistic knowledge and privacy**. Navigating wicked problems requires much information. Often this may require collating data from multiple databases, organisations, and platforms. Collaborative governance requires not only an awareness of the strengths and limitations of digital data, their veracity, but we need to be alert to the how drawing together such data may create risks to personal privacy and security. **Privacy-by-design** is an important dimension of this principle. It means that there should be some checks and balances on the often-unlimited urge to get as much data as possible from as many sources as possible. Practically, this also means avoiding large databases – that act as honey pots for hackers and thus a cyber security nightmare – and instead having distributed databases inter-connecting shared data as and when needed.

4.3 Organisation and Context

The administrative organisation and context of the operation of digital technology is the third domain of design principles.

The first of these is: **Understand the context and the domain of your problem**. This is arguably the starting point in designing digital technology for collaborative governance. If you do not understand the (wicked) problem, then how can a solution be found. This might sound obvious, but when digital technologies are often touted as solutions and arguments that using big data to train machine learning does not require knowledge of the content [33], then this obvious observation is often overlooked.

The twelfth design principle is **digital design and development requires multidisciplinary teams**. This is firstly because digital design is not simply about technology design, but about the design of a dynamic policy-administrative-organisational infrastructure. Secondly, while computer professionals, systems consultants and data analysts are crucial they typically do not have insights into the systems their digital technology operates in nor of the ways in which their data and algorithms structure human interaction and knowledge. Ethical, inclusive, and responsible digital administration requires insights from administrators, citizen-users especially those who are marginalised such as

by disability. Ethical AI or similar principles and frameworks cannot simply be enacted [34] but must be interpreted and purposefully built by people with diverse knowledge and experiences.

The thirteenth design principle is that **to maintain administrative principles digital systems require legal analysis and overarching governance settings.** As digital technologies can redefine administrative processes and even become *de facto* policy through the administrative decisions they enact, responsible administration requires careful legal analysis and possibly development. Digital technology creates challenges to the achievement of administrative justice [35]. Indeed, a key question to assess is whether an automated decision is a legal decision under law; that is, is the computer an authorised state actor. Another question to answer is does the digital process involve a change of administrative principles. For example, Australia's Robodebt reversed the onus of proof of debt from the state being required to make the case, to the citizen required to prove otherwise [13].

This principle is particularly important as already explained – technology is not neutral – is it not a simple translation of human administration practices to digit automated practices. Indeed, digital technologies are often deployed for administrative practices which do not amount to a legal administrative decision. For example, assessing citizens as high or low risk of long-term employment for receipt of employment services, or highlighting citizens of high compliance risks for further investigation or the formatting of interfaces which may exclude users, do not constitute a legal decision and therefore not typically appealable. Accordingly, these forms of automation can operate in a subterranean legal space.

A corollary of this design principle is the need to review and revise administrative justice processes and procedures to ensure administrative justice provisions and rights are unchanged [35].

Identify the appropriate relationship between humans and automation in administrative decision making, is design principle fourteen. This idea is called human-in-the-loop automation. There is a continuum between fully human administration to fully automated administration and the right balance will be context specific. In wholly or highly automated administration, it is still necessary to identify who is (legally) responsible and accountable when the decision is incorrect or creates damage. This is particularly important when governments use external providers of digital technology whereby contractual or commercial-in-confidence provisions become a barrier for both administrators and citizens in understanding automated decisions, assessing potential bias or discrimination in the algorithms, and addressing any flaws in the computer system.

Identifying an appropriate balance between humans and automation is also needed in the drive to 'automate everything'. Automation elides administrative discretion. Computer algorithms require clarity. Thus, while automation may be useful in most cases, due to the complexity of human experience, there will some circumstances when human administrators are necessary. One approach is to use automation to identify when it works well and when cases require human intervention and consideration.

The final design principle is to **create an implementation and transition team including top-level leadership and resourcing**. Organizational and digital change is

difficult and will not be successful without a comprehensive organizational strategy involving key stakeholders reflecting citizens and end users. Government agencies typically have Chief Information Officers (CIOs) or Chief Technology Officers (CTOs) who act as champions, but they must think of what is being undertaken as an organizational strategy, not a digital strategy.

5 Conclusion

Digital technology is making rapid advances in government administration. It is creating new opportunities for politicians, policy makers and administrators to achieve their objectives and enhance services. Digital technologies can be a key part of addressing complex and intractable policy problems requiring collaborative governance. However, digital technology is not without its shortfalls, as the three case studies in this paper have illustrated.

In developing digital technologies for enhancing collaborative governance, administrators need to be mindful of the fifteen design principles outlines in this paper. There are no easy solutions but working through carefully with a critical awareness of the potential role digital technologies can play in collaborative governance is needed for addressing our contemporary and future challenges.

Acknowledgment. This research was conducted as part of the Australian Research Council's Centre of Excellence for Automated Decision Making and Society, though was not funded by the Centre.

References

1. Rittel, H. W., & Webber, M. M. (1973). Dilemmas in a general theory of planning. Policy sciences, 4(2), 155–169.
2. Head, B.W.: Wicked problems in public policy: understanding and responding to complex challenges, Palgrave MacMillan, Cham (2022).
3. Stivers, R.: Technology as magic: The triumph of the irrational. Continuum, New York (2001).
4. Henman, P.W.F.: Digital Social Policy: Past, Present, Future. Journal of Social Policy, 51(3), 535-550, (2022).
5. Perri, P.: e-Governance: Styles of political judgment in the information age polity. Springer (2004).
6. Bovens M., Zouridis S.: From street-level to system-level bureaucracies: how information and communication technology is transforming administrative discretion and constitutional control. Public Administration Review, 62(2), 174–184,(2002).
7. MacArthur, B. D., Dorobantu, C. L., Margetts, H. Z.: Resilient government requires data science reform. Nature Human Behaviour, 6(8), 1035–1037 (2022).
8. Scott, J.C.: Seeing like a state. Yale University Press, (2008).
9. Banerjee, S.: Aadhaar: Digital inclusion and public services in India, World Development Report, pp. 81–92, (2016).
10. Jain, M.: The Aadhaar Card: Cybersecurity Issues with India's Biometric Experiment. The Henry M. Jackson School of International Studies, University of Washington, (2019).

11. Singh R., Jackson, S.: Seeing Like an Infrastructure: Low-resolution Citizens and the Aadhaar Identification Project. In: Proceedings of the ACM on Human-Computer Interac-tion, vol. 5 (CSCW2), 1–26, (2021).
12. Henman, P.: The computer says' DEBT': Towards a critical sociology of algorithms and al-gorithmic governance. Data for Policy Conference, London (2017).
13. Carney, T.: Robo-debt illegality: The seven veils of failed guarantees of the rule of law?, Al-ternative Law Journal 44(1), 4–10 (2019).
14. Whiteford, P.: Debt by design: The anatomy of a social policy fiasco–Or was it something worse?, Australian Journal of Public Administration 80(2), 340–360 (2021).
15. Chouldechova, A. Benavides-Prado, D. Fialko, O., Vaithianathan, R.: A case study of algo-rithm-assisted decision making in child maltreatment hotline screening decisions, In: Confer-ence on Fairness, Accountability and Transparency, pp. 134–148, PMLR (2018).
16. Dare, T., Gambrill, E.: Ethical analysis: Predictive risk models at call screening for Alleghe-ny County, Alleghany County Analytics (2017).
17. Eubanks, V.: Automating inequality: How high-tech tools profile, police, and punish the poor, St. Martin's Press (2018).
18. Henman, P.: Does computerisation save governments money?, Information infrastructure and Policy, 5, 235–52 (1996).
19. Ashworth, J.: Rishi Sunak says technical problems stopped him raising benefits more, The Guardian (2022).
20. Bucher, T.: If... then: Algorithmic power and politics. Oxford University Press, Oxford (2018).
21. Winner, L.: Do Artifacts Have Politics?, Daedalus 109(1) 121–136, (1980).
22. Virilio, P.: Politics of the very worst: An interview with Philippe Petit, Semiotext(e) (1999).
23. Taulli, T., Oni, M.: Artificial intelligence basics, Berkeley, Apress, CA (2019).
24. Heidegger, M.: Being and time. Harper & Row, New York (1962)
25. Fountain, J.E.: Building the virtual state: Information technology and institutional change. Brookings Institution Press (2004).
26. Kane, G.C., Palmer, D., Phillips, A.N., Kiron, D., Buckley, N.: Strategy, not technology, drives digital transformation, MIT Sloan Management Review and Deloitte University Press (2015).
27. Mayer-Schönberger, V., Cukier, K.: Big data: A revolution that will transform how we live, work, and think, Houghton Mifflin Harcourt (2013).
28. Boyd, D., Crawford, K.: Critical questions for big data: Provocations for a cultural, techno-logical, and scholarly phenomenon. Information, Communication & Society 15(5), 662–679 (2012).
29. Uliasz, R.: Seeing like an algorithm: operative images and emergent subjects. AI & Society 16, 1–9 (2020).
30. Henman, P.: The role of computers in texturing micro-social environments. The Australian and New Zealand Journal of Sociology 31(1), 49–63 (1995).
31. Weizenbaum, J.: Computer power and human reason: From judgment to calculation. W. H. Freeman & Co (1976).
32. Hogan, B.: The presentation of self in the age of social media: Distinguishing performances and exhibitions online. Bulletin of Science, Technology & Society 30(6), 377–386 (2010).
33. Mazzocchi, F.: Could Big Data be the end of theory in science? A few remarks on the epis-temology of data-driven science. EMBO reports 16(10), 1250–5 (2015).
34. Fest, I., Wieringa, M., Wagner, B.: Paper vs. practice: How legal and ethical frameworks in-fluence public sector data professionals in the Netherlands. Patterns, 3, 1–14 (2022).
35. Henman, P.W.F.: Administrative justice in a digital world: Challenges and solutions. In: M. Hertogh, R. Kirkham, R. Thomas and J. Tomlinson (ed.) The Oxford Handbook of Admin-istrative Justice, Oxford: Oxford Universtiy Press pp. 459–480 (2021).