# Big Data: Consulting the Public in Public Policy

Högberg, Charlotte

*Published in:*
Encyclopedia of Public Policy

*DOI:*
10.1007/978-3-030-90434-0_60-1

2023

*Document Version:*
Peer reviewed version (aka post-print)

This is the submitted revised manuscript. For citation, please refer to the final published version:

Högberg, C. (2023). Big Data: Consulting the Public in Public Policy. In: van Gerven, M., Rothmayr Allison, C., Schubert, K. (eds) Encyclopedia of Public Policy. Springer, Cham. https://doi.org/10.1007/978-3-030-90434-0_60-1

# Big Data – Consulting the Public in Public Policy

**By:** Charlotte Högberg, Technology and Society, Faculty of Engineering, Lund University

## Keywords

Datafication, big data, algorithms, AI, data collection, data practice, data subject, data citizen, public consulting, public policy, public participation

## Central Definition/Definition(s)

The ability of the government and public sector to *consult* and *know* the public through datafied means is an idea that has been amplified by the development and discourse of *big data*, a concept referring to the collection of large amounts of digital data and expansion of analytical methods. While the governments and public sector are conducting data collection and consulting in their own capacity, it is however commonly part of arrangements with private technology contractors. Digitalization and data analyses are argued to pose new possibilities for data input by and about the public. It includes data deriving from explicit public consultation, co-production with members of the public, public input by digital communication technologies, citizen-generated or collected data, but also implicit data and knowledge extraction from the public. These activities are undertaken to facilitate collection and analysis of data to inform future directions for public policy. Further, the public sector is itself to a great extent undergoing *datafication*, meaning that it is, in addition to digitalization of services, increasingly involved in practices of generating, collecting, analyzing and using digital data about citizens and their conditions. This is incorporated into the making, and execution, of public policy. In the midst of this development, we have both the people that the data is about, labelled as *data subjects*, as well as what has been called the "*data citizen*" – a term referring to humans, and their civic capacity, as seen through digital traces as well as their practices involving digital data.

## Main text

### Historical/generic aspects

The datafication of public sector and public policy is made possible by digitalization, with the aims of streamlining administration and welfare provision. In the ongoing datafication of public sector and policy-making, humans, actions and phenomena are increasingly quantified and turned into digital data (Andreassen et al., 2021). Moreover, the goals for data analysis in public sector tend to focus on the possibilities of prediction, and policy creation capacities, to support performance (Ingrams, 2019: 131). This development is however based on ideological assumptions regarding

the contributions of data analysis. That is, the idea that it will result in increased efficiency, and that data and quantification processes are neutral technologies as well as able to better provide for evidence-based decision-making in public policy. Analyses of historical data are used, or expected to be used, to predict future needs and behaviors, impacting public policy. However, algorithms, models and systems enabling this work, are to large extent seen as value-free and unbiased, and by that, as a safer road to equality of public policy, in comparison to processes in which the human involvement is larger (Andreassen et al., 2021). The datafication of public administration and policy, is also described as entailing processes that exacerbate state control and intensify current power structures (Broomfield and Reutter, 2022: 9-10).

Datafication within public policy has followed a period of discussions about the perceived potential of big data collection and analysis. Moreover, the development of big data has also been argued to have a potential to greatly increase government–citizen understanding, by utilization of the collection and analysis of data about the citizens, their conditions and needs, and by that also improve public policies and services. This potential, argues Ingrams, leads to new dilemmas of public value trade-offs, between matters such as equal treatment and personalization, privacy and transparency, expert- and citizen-informed decision-making (Ingrams, 2019: 129-130). The question is whether it is possible, and if so how, to utilize technological processes of data collection and analysis in a way that is in line with public values. Datafied public policy could risk resulting in "goal displacement", hampering possibilities for external political actors to impact public policy, and assigning disproportionally emphasis on the expert-input in public policy. In large, the datafied public policy is argued to provide opportunities for greater knowledge of, and responsiveness to, public values and needs, but also possible tensions between technocratic values and citizen values (Ingrams, 2019: 130).

In the datafication of public policy, there are different sorts of opportunities for public data input, that can be used to support public policy. It can be explicit public consultation, by means such as digital platforms, open digital forms and forums, facilitating the collection of opinions or needs of the public. Furthermore, it can incorporate public input provided through other types of digital technologies for communication. This includes social media and digital public response to public policy directions, as platformisation has enabled turning communication into data (Andreassen et al., 2021: 214). Digital means of communication can be used as tools for co-creation and co-production of public policy, and by the use of big data collection and analyses, other forms of political representation could be made possible (Lember et al., 2019; Ingrams, 2019).

In addition, public data input is further conducted by implicit data and knowledge extraction from the public. That is, the collection of data about the members of the public, for analysis and use for public policy purposes. It could be big (as in big data) or smaller targeted collections, analyzed for certain problem areas. The subjects of this data collection, the humans that constitutes the public, are increasingly regarded by their position as *data subjects*. This term can be found in regulatory frameworks as well as in research and public discourse, commonly referring to the rights of data subjects in the digital realm, both as users of digital systems and as passive subjects of data collection. While much data is generated by documentation about members of the public, data is also to large extent user-generated, through different kinds of interactions and platforms. In addition, data collection can also be conducted by for example crowdsourcing of data on certain

topics, on web platforms (Gabrys, 2019: 260). However, it can also involve public input by other types of user-generated or collected data.

The term *"data citizen"* is referring to humans, and their civic capacity, as seen through digital traces about them but foremost through their practices involving digital data. For the purpose of bringing forth new types of evidence, citizens are generating, collecting, analyzing and communicating data. One example is Data for Black Lives, aiming to, by alternative data collection and analyses, being able to create new narratives as well as reveal systemic racism (Gabrys, 2019: 250). By those kinds or practices, data citizens are identified by enactments that are forming certain relations and communities, as well as political subjects, instead of by belongings to nation-states or groups. In that way, data citizens are interpreted as "techno-political actors through their data practices" (Gabrys, 2019: 255). Moreover, through datafication, it is argued that data subjects are exaggeratedly identified by belonging to pre-defined categories. Hence, the actions and conditions of previous members of the categorical belonging come to play a larger role, than actions or conditions tied to the present-day individual (Broomfield and Reutter, 2022: 212).

**Functional aspects**

Data analyses could be made to identify public values and needs, however they can be performed in different ways, by various means and goals (Ingrams, 2019). From a historical perspective, the modern state has during a long time been interrelated with the use of statistical data. The development of digitalization aimed at turning all analogue processes into binary code to enable computer processing, and further, the aim of datafication is to quantify all aspects of life into structured data to enable analysis (Broomfield and Reutter, 2022: 2-3). The complexity of this endeavor varies.

The input of digital data about matters for public policy, can be made through digital manual work or automated processes, and further facilitate manual or automated analyses, by use of artificial intelligence technologies such as machine learning, but also more rudimentary decision trees (Andreassen et al., 2021). The public sector can use data collection for public engagement and public participation, as a way to perform public consultancy by digital and datafied means. It requires digital tools enabling data collection, data analysis and use, and these are often part of power structures that typically are working against public social policy (Andreassen et al., 2021: 212). Much of data collection and digital public consulting is also reactive citizen engagement, rather than co-creation, depending on openness of information and modes of policy-making and project ideals. Ingrams (2019: 142) argues that efforts to support public participation through data involvement for matters as shared problem solving, have not yet seen much realization, due to lacking institutional capacity.

Public data input in terms of digital data, is connected to many different types of technologies: such as sensing technologies, communication technologies, processing technologies, actuating technologies (Lember et al., 2019). Studies of data citizens have shown how data is generated through technologies such as wearables and social media, but it can also be citizen sensing technologies for generating data of matters as air pollution and other conditions, pertaining to the gathering of evidence of the lived experience of citizens (Gabrys, 2019: 255-257). The problematic aspects of datafied public input, is interrelated to general issues identified as brought about or

exacerbated by datafied public policy. These include problems of relations between public and private sector, risks of decreased transparency, lack of diversity, and discriminatory bias in data or as a derived consequence of algorithmic data analysis.

Further, datafication is put forward as a matter of concern due to humans not being regarded in terms of individuals and their personal situation, but increasingly as members of categorical belongings in automated processing of quantified data points (Andreassen et al., 2021). Increased digitalization of public services, and automation of decision-making, further risk decreasing and diminishing dialogue and information from the public, that could be important for policy-making. However, some groups could be benefitting from this development, such as those with high digital literacy, while others could be less able to impact their data input for public policy. The same argument can be made about the concept of data citizens. What sort of political subjectivity that people are enacting through data practices, is not clear. The new modes of producing evidence works for some, but not for all. Further, the data still has to be deemed of relevance and be used to make a difference in public policy. Claims of accuracy and legitimization of data in the name expertise, can still impede the impact of public participatory data action (Gabrys, 2019). With regards to the matter of data subjects, it is also important to keep in mind who the data is about. Exclusion from data could have negative impacts of non-representation, yet several studies point to the tendency of vulnerable or marginalized group being to large extent targeted by data collection and surveillance. On a meta level, one can also question how involved citizens are in the datafication process of public policy (Broomfield and Reutter, 2022). Despite the participatory turn, in the datafied public sector, citizens seem to be positioned as passive stakeholders. Studies find that regarding policy directions of datafication, the civil society is rarely directly included, while actors from the private sector are regarded important partners and co-creators (Broomfield and Reutter, 2022: 10).

**Different meanings/definitions/use**

There is a wide array of practices that could be argued to be public data input to inform public policy. There are many adjacent concepts and the terminology is not stable. Data participatory practices could be labelled as consultation, co-production, co-creation, collaboration, communication, crowd sourcing, data citizenship and many other terms (Lember et al., 2019; Gabrys, 2019). Many of the practices can also be said to both work in favor of empowerment, and to challenge it in other aspects. (Big) data collection can support expert-informed policy-making, and place a greater concentration of power onto expert knowledge of domain experts, but also data analytics professionals. Yet, it is also argued to have the potential of bringing greater impact to citizen-informed policy-making enabled through assumed greater knowledge and possibilities of participation (Ingrams, 2019: 130).

Moreover, one case of multiple definitions, with possible political implications, is that of data citizen. The term data citizen is used in various ways within different contexts. In the tech industry, it is referring to someone that with great ease is dealing with data collection and analyzing techniques. The citizen part is added for its seemingly democratic and participatory value. In practice, these values speak perhaps more to how data citizen as a term is used within research and activist communities, even though it is still unclear how people are positioned as political actors through data practice (Gabrys, 2019: 255). The data collected within the communities of data

citizens, are however expected to be able for the people to analyze. Meaning that they have, in some capacity, access to the data and be able to analyze and communicate it. This commonly separates it from crowdsourcing of data, where data is being collected by the public but data is held with restrictive access (Gabrys, 2019: 260). Yet, data citizen is often also used to refer to the data gathering through wearables or other types of technologies, by proprietary collection, and without clear political intention and democratic involvement. Gabrys (2019: 255) argues that depending on the conditions for data collection, to participate can advance de-democratization as well. By this notion, the multiple ontologies of data citizens that are in play, show that how the term is materialized in practice, can constitute very different degrees of empowerment and political subjectivity.

**National differences**

The national differences of public consultancy, public data input, co-creation, data subjects and citizens, are connected to the specific policy directions and values pertained to datafication of the public sector, and the interrelated regulatory frameworks of personal data collection. There are different principles of public sector collection of data and sharing of data. Moreover, in some countries, such as the United States, the public sector depends to a greater extent on private companies' data collection in comparison to Europe (Ingrams, 2019). From a global perspective, the expectations regarding democratic involvement of citizens are also varied. In some settings, datafication is motivated in the name of improving the functions of the welfare state, such as in the Nordic countries (Broomfield and Reutter, 2022; Andreassen et al., 2021). Datafication of public policy is however also developing in comparably very different contexts, with another trajectory and explicit agenda in place, such as the cases of social credit scoring. Still, communities of "data citizens" taking collective action to create other stories of data for the public, can be found in many parts of the world, for example by the oil and gas wells in Pennsylvania, and in proximity of Fukushima, monitoring radiation, as well as in air polluted areas of London (Gabrys, 2019). Moreover, the regulatory frameworks to protect citizens in relation to data also impact the development in different jurisdictions. Yet, in line with what Gabrys (2019: 264) argues, the rights controlling data collection, tracking, to be forgotten, transparency of data, and similar matters, do to little extent tell us about the current formations of conditions for the enacting of rights of political subjects, and modes of engagement, in datafied societies, and what kinds of worlds that are in the making by those processes.

## Cross references

- Big Data – Automated Decision-Making in Public Policy
- Evidence-based Policy Making
- Co-production in Public Policy
- Public-Private-Partnership in Public Policy

## References

Andreassen R, Kaun A and Nikunen K (2021) Fostering the data welfare state: A Nordic perspective on datafication. *Nordicom Review* 42(2): 207-223.

Broomfield H and Reutter L (2022) In search of the citizen in the datafication of public administration. *Big Data & Society* 9(1): 20539517221089302.

Gabrys J (2019) Data citizens – How to reinvent rights. In: Didier B, Isin E and Ruppert E (eds) *Data Politics: Worlds, Subjects, Rights*. London: Routledge, pp.248–266.

Ingrams A (2019) Public Values in the Age of Big Data: A Public Information Perspective. *Policy & Internet* 11(2): 128-148.

Lember V, Brandsen T and Tõnurist P (2019) The potential impacts of digital technologies on co-production and co-creation. *Public Management Review* 21(11): 1665-1686.