

## Report review

---

# Data scores as governance

Dencik, L., Hintz, A., Redden, J., & Warne, H. (2018). *Data Scores as Governance: Investigating uses of citizen scoring in public services*. (Data Justice Lab: Cardiff). The report is freely available from: <https://datajustice.files.wordpress.com/2018/12/data-scores-as-governance-project-report2.pdf>

The Data Justice Lab's report, '*Data scores as governance: investigating uses of citizen scoring in public services*' by Dencik et al. (2018) examines how data analytical practices are used throughout the UK's public sector. As the title suggests, the authors pay special attention to the use of data-driven scores which prioritize, classify, associate, predict, and categorize citizens. Based on a mixed method approach, employing desk research, Freedom of Information requests, stakeholder workshops, and interviews (with representatives of both the practitioner community and the civil society), the report reviews the usages of data analytics across various local governments and other authorities such as the police. It describes the prominent software tools, and provides an exploratory investigation of how stakeholders and civil society perceive the development. This report has been accompanied by a 'Data Scores Investigation Tool' ([www.data-scores.org](http://www.data-scores.org)) which encompasses scraped government documents referring to data practices.

The report itself can roughly be divided into five parts. The first part deals with the analysis of Freedom of Information (FOI) requests to the local authorities concerning their use of data analytics. Secondly, the report presents case studies of particular systems (e.g. the Kent Integrated Dataset which is employed for population health planning), and dominant companies in the field of public sector data analytics (e.g. Experian). Thirdly, the study reports back from four stakeholder workshops which (a) helped scope the research, (b) tested the Data Scores Investigation Tool, (c) investigated ways to enhance citizen participation with regards to data scores, and (d) presented the preliminary findings. Fourthly, the report describes the interviews that the authors of the report conducted with (a) practitioners and (b) civil society. The fifth, and final, section of the report deals with the compilation of government documents in an online tool.

This review of this report will first address the findings with regards to the use of data scores (FOI and case studies). Secondly, it will discuss the perspectives on that usage of data scores (interviews and scoping workshop). Finally, Data Justice Lab's online tool will be evaluated.

In order to investigate the usage of data scores in local governments and other kinds of authorities, Data Justice Lab sought to create a comprehensive overview of the situation of the entire UK by sending out an impressive 423 FOI requests to various local authorities about their use of data analytics (response rate of 24 per cent). 20 of the requests were targeted to specific systems, which formed the body from which eventually the six system case studies were selected. The remaining 403 requests were phrased in a more generic manner in order to allow for the diversity of a great range of systems. The responses to the FOI requests pointed to a lack of 'standard practices or common approaches with regards to how data is or should be shared and used by local authorities and partner agencies' (p. 3).

However, the unspecific request came with some drawbacks as some of the requests (six per cent) were declined by the responding authorities due to imprecise requests (not understanding what data analytics

meant), and/or collecting the material would be too costly. Moreover, there was a non-response rate of 11 per cent and according to the local authorities, in 55 per cent of the cases, there was no available information about the systems in operation, or no systems at all. Consequently, it is difficult to seriously evaluate the use of the FOI request as a data collection method and the report does not discriminate between different categories (e.g. no information received/no system present/request not acknowledged). Such conflation of categories sadly obscures part of the investigation's goal. Nevertheless, the report shares many valuable lessons, including types of systems and brands used in the public sector.

These systems and brands are discussed more in depth in the case study section (six systems for data analytics, profile sketches of one specific software program, and the operations of three dominant companies). The case studies are very helpful for understanding the context in which systems are used, including the respective organizations' rationale for the development and/or implementation of the systems, and their different governance approaches to data systems. While the City Council of Bristol developed its 'Integrated Analytical Hub' in-house, thereby securing independence from third parties, and maintaining control of the workings, the Manchester City Council purchased an off the shelf product from IBM for their Research and Intelligence Database. In this perspective the Camden council represents a compromise between these two extremes. Although using a system supplied by IBM, the data model they are running is accessible to the Council and the data can be adjusted.

In order to shed light on the civil society perspective on the use and impact of data scores, the Data Justice Lab interviewed 10 members of different civil society groups such as 'Big Brother Watch' and 'Open Rights Group'. In addition, the authors conducted a scoping workshop with a combination of different stakeholders to 'a) explore the state of government uses of data analytics, b) investigate challenges and opportunities, and c) offer a space for dialogue between different stakeholders' (p. 110). The civil society group members expressed both benefits and hazards of using data analytics in the public sector. The negative effects comprised 'the extent of data collection and sharing, the potential for bias and discrimination in decision-making, the possibility of targeting, stigma and stereotyping of particular groups, lack of transparency, public knowledge, consent, and oversight, and the limits of regulation to address the overall political context of uses of data systems in public services' (p. 102). The supplementary workshop highlighted tensions on themes such as public-private partnerships, public engagement, accountability, and privacy (p. 111).

The reservations to data analytics articulated by the stakeholders often seem to be at odds with the objectives behind the systems implemented in the public sector. To illustrate this one may take the example of 'Bristol's Integrated Analytical Hub' which sole purpose is to make 'targeted risk assessments'. The successful operation of this systems requires massive input of data in order to provide useful information, and furthermore that this data is accessible across a number of organizations. This is something that naturally clashes with the views of many of the civil society groups concerned about both targeting, and data sharing. Moreover, this only scratches the surface of potential complications, as the model itself, although managed by the Bristol City Council, is often hidden from the public eye. Furthermore, it is particularly difficult to make the 'systems' understand, as a data scientist working on one of the project argues, 'what a family actually is and how to group people together and understanding their needs' (p. 31). This gives away a multitude of problems with stereotyping, stigmas, and biases, as Virginia Eubanks (2018) describes happened in an US context in which risk scores were used to profile child abuse.

The interviews with the 17 practitioners across the six studied systems also illuminated tensions and clashes of professional values within the public sector organizations (p. 34). Moreover, developers and data scientists worry that the end users may interpret the scores incorrectly. As one data scientist puts it: 'I think there's a risk that once something's in a system and somebody sees it and it's got a number next

to it, they think it must be right because the computer's told me that and then they just forget all of their professional knowledge and judgement and say the computer says this' (ibid.). Hence, tensions seem to exist between the developers of the system/model, the users of the system (or the *expected* users of the systems), and the civil society perspectives.

However, in some instances, the report falls short of a deeper, and more granulated analysis of the public sector's adoption of data analytics. While the team behind the report was able to speak with practitioners from all the different stages of the implementations in some instances, this does not seem to hold true for all the cases. Consequently, there seems to be lot of scope for future systematic inquiries, not by only filling the current gaps, but also studying the use of data score tools by civil servants, and their impact on citizens. At present, the report does acknowledge the importance of professional discretions of the end-users, but choses to place emphasis on either the practitioners (e.g. developers, data analysts, project managers) or civil society, with little attention to the end-user.

Perhaps the most intriguing part of Data Justice Lab's project is the launch of its Data Scores Investigation Tool ([www.data-scores.org](http://www.data-scores.org)). This tool (website) serves as a focal point for documents relating to data and algorithmic systems employed by public sector organizations in the UK. At the time of writing, over 5,300 documents have so far been added to the database. These documents are derived from a variety of sources including scraped government websites, newspapers, and FOI requests.

Even though the potential use cases are ample and diverging, the tool does come with some limitations. While the tool allows for searching through the compiled documents rather easily, the search function, at present, is limited. For instance, regular expressions do not seem to be supported (e.g. ['data score\*'] does not return hits for ['data scores']). Other functionalities, such as search conditions (e.g. ['data score' OR 'algorithmic risk assessment']) cannot be used either. Both issues disallow the user from employing sophisticated query designs. A last hindrance is the impossibility of easily downloading a selection of documents, which is not insurmountable, but does slow down potential investigations drawing on a large amount of these texts.

Despite some technical challenges, the tool still provides both researchers and the broader public with a very valuable open database of algorithmic systems employed within a multitude of organizations. The collection of such documents fits within a larger, international, trend of openness, in which FOI requests are made accessible (e.g. Open State Foundation, 2019); open data starts to become the norm (e.g. U.S. General Services Administration, 2018); and information resources are bundled for a general audience, as in the case with 'Algorithm Tips' . (Computational Journalism Lab, 2019) from which Data Justice Lab drew their inspiration. Such focal points of information are greatly beneficial to both the academic and journalistic communities, as they assist in commencing new lines of inquiries. The added value lies precisely within that bundling and searchability of information, thus lowering the technical entry barriers for investigation on data and algorithmic systems within the UK public sector organizations.

Data Justice Lab's report on data scores does what it set out to do. It provides a comprehensive overview of the UK public sector's use of data analytics in general, while offsetting this with more in depth case-studies and a civil society perspective. Combining these perspectives, the empirical studies touch upon many interesting observations and future potential lines of inquiry. While the presentation of some of the report's elements could be stated a bit clearer, or be developed further, the report and its accompanying investigation tool is very valuable. One can only hope that the Data Justice Lab's inquiries have only just begun.

Maranke Wieringa, Department of Media and Cultural Studies (Datafied Society), Utrecht University, e-mail: [m.a.wieringa@uu.nl](mailto:m.a.wieringa@uu.nl)

## References

- Computational Journalism Lab. (2019). Algorithm Tips. Retrieved January 23, 2019, from <http://algorithmtips.org/>.
- Data Justice Lab. (2019). Data Scores. Retrieved January 23, 2019, from <https://www.data-scores.org/>.
- Dencik, L., Hintz, A., Redden, J., & Warne, H. (2018). *Data Scores as Governance: Investigating uses of citizen scoring in public services*. Cardiff. Retrieved from <https://datajustice.files.wordpress.com/2018/12/data-scores-as-governance-project-report2.pdf>.
- Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.
- Open State Foundation. (2019). Open WOB. Retrieved January 23, 2019, from <https://www.openwob.nl/>.
- U.S. General Services Administration. (2018). Data.gov. Retrieved July 4, 2018, from <https://www.data.gov/>.