Data Scores as Governance: Mapping and Analysing Changing Practices in the UK

Prepared for the interdisciplinary symposium "Super-Scoring? Data-driven societal technologies in China and Western-style democracies as a new challenge for education." Cologne, Germany; October 11, 2019. The essay can be downloaded here as a PDF-File.

By Joanna Redden, Lina Dencik, Arne Hintz and Harry Warne

This essay details research undertaken by a Data Justice Lab research team. The aim of the Data Scores as Governance project is to map and analyse local government uses of data analytics in the UK, with a particular focus on investigating uses of predictive scoring systems. Our multi-method investigation led to: 1) a comprehensive list and map of data analytics systems across local authorities, 2) a research report that details concrete examples of the different types of analytics systems being used as well as a survey of civil society concerns and 3) an interactive online tool to facilitate greater research and debate.

Introduction

Our research is motivated by a recognition that we know governments at all levels are implementing predictive scoring systems and advanced analytics to make decisions that affect public services, but we know too little about where and how these systems are being used. This means that we also struggle to fully appreciate the larger implications that come with the implementation of these systems and also risk viewing all systems as the same. Government agencies are drawn to predictive scoring systems because they promise to help them become more efficient and better target services in an age of cuts and constraints. Research into the use of predictive scoring systems raises concerns about how they may limit people's access to services and lead to greater inequality and discrimination (Eubanks 2018, Barocas and Selbst 2016, Angwin et al. 2016, Gillingham and Graham 2017, Keddell 2015, O'Neil 2016).

Building on the work of others, we argue that we need to identify where and how scoring systems are being used to better understand changing data systems as part of their wider contexts and to also recognize actor agency and attend to situated practices (Kennedy 2016, Couldry and Powell 2014, Dencik 2019). We find a range of experimentation and uses of data systems across the UK, and scoring systems in particular being introduced in areas of child welfare, fraud detection, policing, and public safety and transport. Our findings indicate that these systems are heterogeneous and contingent upon situated and contextual factors. Broadly, we find that the turn to predictive scoring systems as well as other systems that rely on a vast collection and sharing of data is driven by an austerity context in which local authorities are under great pressure to do more with less. Our research raises questions about the impact of these systems on resource allocation, frontline service delivery, the

longer term impact of data sharing and systems of 'knowing' and responding to issues and people.

Methodology

In our research we systematically mapped the use of 'advanced analytics' across local authorities in the UK while also conducting more micro level research through a focus on six case studies. A detailed discussion of our methodology can be found in our report (Dencik et al. 2018). In brief, our research involved interviews, workshops, freedom of information requests, consulting grey literature and computational methods. We conducted 27 semi-structured interviews with public sector workers (17) and civil society groups (10). In these interviews we sought to better understand the benefits, challenges and concerns identified as linked to these systems. We submitted 423 Freedom of Information request to local authorities and agencies, twenty of these were targeted requests and the other 403 were general requests submitted via mySociety's WhatDoTheyKnow online service. We held workshops to bring practitioners from different sectors together and consider the range of practices and debate. Finally, we used computational methods to construct a Data Scores Investigation tool drawing on the methodology of the Algorithm Tips project. This involved the use of search engines to scrape documents from UK government sites (gov.uk, nhs.uk, police.uk, mod.uk and sch.uk) and media sites based on a list of keywords relating to data analytics and algorithmic decision-making.

Discussion

Our research identified a range of data analytics systems in use and that uses of data analytics across local authorities are distinct. We found uses of data analytics and visualization tools: to identify connections in an effort to better understand families and individuals; uses of predictive scoring systems to anticipate risk for individuals and families; and uses of data analytics for population level analytics to anticipate current and future needs. Despite the diversity of applications, all of these systems require linking up multiple datasets, which includes accessing highly sensitive and personal data. These large combined datasets are commonly referred to as data warehouses and data lakes. A number of local authorities are developing these data warehouses, but not using scoring systems. We found that predictive scoring systems are being used in child welfare, policing, fraud detection, public safety and transport amongst other areas. Some councils are developing their own in-house systems, others are working with private companies.

In some cases the linking up of data systems is done to make it easier for frontline staff to share information and produce more comprehensive individual and social network profiles and in other cases to enable automated risk analyses to produce alerts when a certain threshold has been crossed. A recurring theme is that the austerity context and the funding constraints faced by local authorities is one of the main reasons for introducing or enhancing data analytics systems. Local authorities are responding to cuts by using data to try and

better target resources. Full details of the systems referenced below are available in our report.

A focus on uses of data analytics in child welfare provides an illustrative example of the different kinds of systems being used. It is important that we understand the use of scoring systems as part of this wider context of data system practices within public services because an awareness of the diversity of applications and a grounded understanding of how they work is needed to better inform debate and policy responses. All of the data systems we identified being used in child welfare were developed in response to the government's 'Troubled Families' Programme, which is itself controversial. The Bristol Integrated Analytical Hub is developed in-house to make use of a database that consolidates 35 social issue datasets about 54,000 families. Initially the Hub was created to provide a 'holistic understanding' of families (manager). After the warehouse was developed the team in Bristol started looking into ways to use the data to predict future need and created a model to predict child sexual exploitation. This trajectory demonstrates that the collection and combination of datasets, once done, can stimulate interest in the use of predictive scoring systems. In Bristol, there was a deliberate intention to develop the system inhouse to maintain control of the system and the data. The London Borough of Hackney is another local authority that has been trying to use scoring systems to anticipate risk of harm to children and families. They contracted Xantura to provide this system. The predictive model in this case combines data from multiple agencies. The system produces monthly risk profiles in the form of a report that is sent to social workers working with those families identified as in most need of intervention (LC 2018). Manchester does not use a predictive scoring system, although they are investigating how they might use predictive analytics going forward. Manchester City Council purchased IBM's iBase system and then modified and developed it according to their needs. The data warehouse they created combines 16 datasets and case workers are able to access data going back five years. The system is used to make case worker access to information about families more efficient and to also identify families that meet the 'Troubled Families' criteria which can lead to more funding for the local authority and is supposed to lead to more support for the families identified.

The need to link data in order to identify families that meet the central government's Troubled Families Programme criteria is a driving force behind these new data practices in child welfare and the shape they have taken. Concerns have been raised about how the programme locates 'troubles' or 'problems' with the family and individual without attending to wider systemic and economic factors (Lambert and Crossley 2017). Further, the need to access greater funds and the work done to do so is a product of an austerity context which has seen local authorities in the UK face major funding cuts from central government. This finding points to the political contingency of these systems and how they can be products of a particular policy context which can then reinforce that context.

The systems analysed have different practices in terms of notifying people that their data is being used and also in terms of seeking consent for the use of this data. This demonstrates a diversity of approaches and opinion within government bodies on this issue.

The differing approaches to consent lead to questions, in some cases, about whether or not people who don't know their data is being used as part of these systems are able to exercise their rights through the GDPR. Across all of our case studies we identified differing levels of transparency and public accountability. We found that the issue of data quality as well as concerns about the accuracy of predictive scoring systems require far more attention. We also identified little effort to encourage citizen engagement and intervention. There is also little assessment of unintended consequences and of the impact of interventions taken on the basis of data-driven scores. Our interviews with civil society organizations indicate that the key concerns for them go beyond questions of transparency, as concerns were raised about targeting and stigmatisation and changes in the way governments come to 'see' and engage with service users. There is a need for nuanced debate that separates out and considers data sharing systems and predictive scoring systems.

Conclusion

Our research points to the need to better understand how the data systems being introduced are changing working practices of frontline professionals and resource allocation and also how these systems, over the longer term, may shift government priorities and the way government agencies come to know and engage with people. For example, we raise concerns about how a focus on capturing and analysing data in relation to individuals may direct attention away from the need to capture data about the influence of positive or insulating factors that can reduce risk such as the presence of extended family networks or an afterschool program. More broadly it may focus responses solely on the individual or household, bypassing societal factors in the creation of social problems. Finally, we raise concerns about how an emphasis on risk assessment may lead to a broader shift in state operations as citizens become viewed less as co-creators of the societies they are part of and more as potential risks needing management (McQuillan 2018). There is a significant disparity between practitioners' and stakeholder groups' perspectives on the nature of challenges that emerge from uses of scoring systems and data analytics in public services more generally demonstrating the need to bring these groups together and expand cross-sector debate and enhance the means for meaningful citizen participation and intervention.

References

Angwin, J., Larson, J., Mattu, S. and Kirchner, Lauren (2016) 'Machine bias', *ProPublica*, 23
         May. Available from: https://www.propublica.org/article/machine-bias-risk-
         assessments-in-criminal-sentencing (Accessed: 2 Sept. 2016).
Barocas, Solon and Selbst, Andrew D. 2016. Big data's Disparate Impact.
         California Law Review. 104: 671–732.
Couldry, Nick and Powell, Allison. 2014. Big Data from the Bottom Up. *Big Data & Society*.
         1(2): 1–5.

Dencik, Lina, Hintz, Arne, Redden, Joanna & Warne, Harry. (2018). *Data Scores as Governance: Investigating Uses of Citizen Scoring in Public Services*. Project Report. Data Justice Lab, URL: https://datajustice.files.wordpress.com/2018/12/data-scores-as-governance-project-report2.pdf [1 March 2019]

Dencik, Lina. 2019. Situating practices in datafication – from above and below. In H. C. Stephansen & E. Treré (Eds.), *Citizen Media and Practice*. London; New York: Routledge.

Eubanks, Virginia..2018. *Automating Inequality*. New York: Macmillan.

Gillingham, Philip & Graham, Timothy. (2017). Big data in social welfare: the development of a critical perspective on social work's latest "electronic turn." *Australian Social Work*, 70:2, 135-147.

Keddell, E. 2015. The ethics of predictive risk modelling in the Aotearoa/New Zealand child Welfare context: Child abuse prevention or neo-liberal tool? *Critical Social Policy*, *35*(1), 69–88. doi:10.1177/0261018314543224

Kennedy Helen. 2016. *Post, Mine, Repeat: Social Media Data Mining Becomes Ordinary*. Basingstoke: Palgrave Macmillan.

Lambert, Michael and Crossley, Stephen (2017) Getting with the (troubled families) programme: a review, *Social Policy and Society*, 16(1): 87-97.

London Councils. 2018. Venture Spotlight: EHPS, available: https://www.londoncouncils.gov.uk/node/31412

McQuillan, Dan. (2018). People's councils for ethical machine learning. *Social Media + Society*: 1-10.

O'Neill, C. (2016) *Weapons of Math Destruction*, New York: Crown.