COGITATIO

Article

Realising the Benefits of Integrated Data for Local Policymaking: Rhetoric versus Reality

Hannah Durrant [1,*], Julie Barnett [1,2] and Emily Suzanne Rempel [1,2]

[1] Institute for Policy Research, University of Bath, Bath, BA2 7AY, UK; E-Mails: h.durrant@bath.ac.uk (H.D.); j.c.barnett@bath.ac.uk (J.B.), e.s.rempel@bath.ac.uk (E.S.R.)
[2] Department of Psychology, University of Bath, Bath, BA2 7AY, UK

* Corresponding author

Abstract
This article presents findings from local government projects to realise the benefits of big data for policy. Through participatory action research with two local statutory authorities in the South West of England, we observed the activities of identifying, integrating and analysing multiple and diverse forms of data, including large administrative datasets, to generate insights on live policy priorities and inform decision-making. We reveal the significance of both data production and policymaking contexts in explaining how big data of this kind can be called upon and enacted in policy processes.

Issue
This article is part of the issue "Big Data Applications in Governance and Policy", edited by Sarah Giest (Leiden University, The Netherlands) and Reuben Ng (National University of Singapore, Singapore).

1. Introduction

The claims made for big data in business contexts are well established (e.g., Mayer-Schönberger & Cukier, 2013). Kitchin (2014b) discusses the powerful sets of discourses that are employed to support the application of big data to realise tangible improvements to business processes, products and profits. These include, but are not limited to, the ability of big data technologies to enhance logistics planning, reduce inefficiencies, understand customer preferences, target products and services to new and existing markets and combat fraud (Kitchin, 2014b, pp. 117–123). More recently, attention has turned to the potential of big data for policymaking settings (e.g., United Kingdom [UK] Parliament, 2015), and the challenges involved in harnessing this potential to realise policy aims and objectives for the public good (Janssen, Konopnicki, Snowdon, & Adegboyega, 2017; Kennedy, Moss, Birchall, & Moshonas, 2015; Malomo & Sena, 2016; Schintler & Kulkarni, 2014). Administrative (open) data is particularly prominent in Poel, Meyer and Schroeder's (2018) analysis of the use of big data in policymaking, being used in two thirds of the 58 such initiatives they identified. Questions have been raised about how and where in the 'policy cycle' big data-derived analysis could feed in (Höchtl, Parycek, & Schöllhammer, 2015), with increasing emphasis being placed on the role that data can play in predicting need and defining policy priorities for the future (Giest, 2017; Malomo & Sena, 2016). This work usefully disaggregates the applications of data, moves beyond rhetoric and opens up thinking about the spaces for data science to inform policymaking.

However, policymaking processes are not straightforward or linear, and there is a need to theorise the social contexts of both data production and policymaking to understand the boundaries and barriers to big data for policy in practice. We set out to reveal the temporally-specific and contingent ways in which data are articulated in the demand for evidence, and discuss how

the practices and preoccupations of policymaking both shape and are being shaped by the promise of data.

The article unfolds as follows. We begin (in section two) by rehearsing the claims that have been made about big data, and that have sought to give this ubiquitous but simultaneously elusive term some definitional clarity. We focus on claims made about the promise of data for policymaking, and problematize assumptions of linear and rational policymaking processes into and through which data science can flow. We rather propose a counter theory of policymaking as struggles over the right to advance ideas about policy; why it is needed, what it should do, for whom, how and to what end (Carmel & Papadopoulos, 2003). We argue that it is in these deeply value-laden and political contexts that data are produced and repurposed, and insights are allowed, or otherwise, to be admitted as a form of evidence.

Section three briefly describes the participatory action research approach adopted in this project and details the partnership and processes by which the project progressed. Section four presents findings and reflections from the project; focusing on the ways in which data is constituted as relevant to policymaking, the terms on which its use is resisted; and the importance of relationships of trust to underpin data processes in practice. We conclude in section five by discussing the significance of the social context of both data generation and policymaking to explain what can actually be done with data in policy settings.

2. Big Data and Policy Making

Conventional attempts to define big data have tended to focus primarily on its characteristics; initially emphasising its volume, variety, and velocity (see Kitchin & McArdle, 2016). A more recent proliferation of characteristics identified with big data (e.g., Uprichard, 2013) has rendered the term more, rather than less, opaque (Kitchin & McArdle, 2016). In an attempt to isolate the most salient qualities of big data, Kitchin (2014b) stresses the distinction between small data sources—based on a population sample, infrequently collected and processed slowly and periodically—and big data that is both exhaustive (n = all) and generated and reported in close to real time. For Kitchin and McArdle (2016) the two most important characteristics of big data sources are exhaustivity and velocity.

Consideration of the sources of data has also been used to ground understanding of what is commonly considered to constitute big data (Connelly, Playford, Gayle, & Dibben, 2016). Data are being generated from a greater variety of sources than ever before (Kitchin, 2014a). Some of this data is what Mayer-Schönberger & Cukier (2013, p. 113) refer to as 'data exhaust'; the by-product of people's digital activities and interactions (e.g., financial transactions and social media activity), repurposed to another end. For Connelly et al. (2016), the 'found' nature of big data, and its ability to be valuably re-

purposed, is a significant feature. They differentiate between data that is 'made' by social scientists to study the social world, and data generated for entirely different purposes yet possessing considerable research utility (see also Cowls & Schroeder, 2015).

The identification of a wide array of characteristics and sources of big data conveys a sense of its ubiquity, but also the extent to which it has defied definitional clarity. Recent scholarship has begun to systematise 'types' of big data according to the types of traits that it possesses (Connelly et al., 2016; Kitchin & McArdle, 2016). Both of these articles identify multiple types or forms of big data. In particular, Connelly et al. (2016) assess the extent to which administrative data, a form of data derived in the process of administering services and systems and commonly held by UK government (nationally and locally) and other public sector bodies, can be considered one form of big data. They argue that it meets the conditions because it is often exhaustive, highly granular, large (as a consequence of being both exhaustive and highly granular), messy and unstructured and, importantly, found and repurposed rather than made (see also Kitchin & McArdle, 2016; Malomo & Sena, 2016).

We would agree with this assessment. In our experience, working with statutory bodies in the South West of England, we found that local government administrative data, particularly when integrated with other forms of demographic, contextual and unstructured data, demonstrated many of the characteristic of big data. Data relating to, for instance, primary and secondary health care, social benefit claims, and the delivery of public services, cover the entire population (i.e., all patients, claimants and service recipients within an administrative boundary). In addition, administrative data are often produced in real time and can be extracted for use at frequencies close to real time. They are granular to the extent that they are individual-level and contain extensive fields; including details of service use, as well as demographic, service process and background information. Granularity is enhanced further where data is linked and integrated, and we found that some datasets contained both structured and unstructured data (e.g., case notes and service user comments and feedback). Most importantly, however, these data were found to be of value to social science research and policymaking, rather than made.

Furthermore, and in line with other scholars that have focused on the benefits of big data for policymaking, we include administrative data as a source of big data—particularly where it is linked and integrated with other data sources—on the grounds of its particular relevance and value for policy (Connelly et al., 2016; Poel et al., 2018). Administrative records provide governments at all levels in the UK with unique access to diverse data generated on the people and communities they serve, and there is a growing literature on the application of these kinds of data in policy settings (Janssen et al., 2017; Malomo & Sena, 2016; Poel et al., 2018).

Current data initiatives are accompanied by powerful rhetoric about the significance of big data for policy, emanating from within policymaking communities. In 2015, the UK Parliament identified harnessing the benefits of big data as a key issue for government(s); describing data as "the new oil" (UK Parliament, 2015) and, just as is the case in business contexts, here too the claims for the possibilities afforded by big data are expansive. Stephan Shakespeare, in his review of Public Sector Information, enthusiastically asserts that "from data we will get the cure for cancer as well as better hospitals; schools that adapt to children's needs making them happier and smarter; better policing and safer homes; and of course jobs" (Shakespeare, 2013, p. 5). Thus, as well as implying a set of characteristics, the term 'Big Data'—coined in the context of a data revolution that government(s) in the UK are keen to capitalise on—is also pervaded by a set of strongly held and asserted beliefs about the purposes to which data can be put and the ends that are envisaged (Kitchin, 2014b; Markus & Topi, 2015). Markus and Topi (2015) contend that definitions should acknowledge big data as more than sets of data with particular characteristics that require novel analytical techniques, and equally recognise the ideas that seek to inspire its use (see also boyd & Crawford, 2012). They argue for viewing big data as, "a cluster or assemblage of data-related ideas, resources and practices" (Markus & Topi, 2015, p. 3).

Optimistic claims for the potential of big data tend to obscure challenges associated with its use. At the most extreme, big data advocates promote a view that data—in great enough volume and when properly interrogated—can "speak for themselves" (Anderson, 2008), and that this may be a welcome step forward for evidence-based policymaking (Mayer-Schönberger & Cukier, 2013). More recently, more critical approaches have questioned the portrayal of data as neutral and data science as objective; raising the politics of data capture and analysis.

A growing critical data studies literature (Iliadis & Russo, 2016) has emphasised that data is generated, curated, processed and interpreted through frameworks that determine what is constituted as data and how it can be translated into information (boyd & Crawford, 2012; Kitchin, 2014a). Such frameworks are inherently political because of what they count and what they leave out, what they make visible and what they render invisible; particularly when being visible and counted is a necessary precondition in qualifying for political, economic and social resources. As Johnson (2015) states: "the ability to make one's group, and one's interests legible to the state, organizations, or other individuals is increasingly determined by where one stands in the data".

Kennedy et al. (2015, p. 175) point to a widespread awareness—particularly among social scientists—of the ways in which data is shaped and given value by the context in which it is produced and the methods by which it is aligned, processed and analysed. Following them, we sought to understand the extent to which data, and the

techniques for extracting meaning from it, came under critical examination in the practices and processes of policymaking. We note that the claims regarding the potential of big data for policymaking are often disconnected from the sets of ideas, resources and practices involved in data application to policy. This article is concerned with understanding the narratives, processes and practices by which data can meaningfully grease the wheels of decision-making in policy settings.

Recent scholarship has sought to identify opportunities for big data insights to inform policymaking, by focusing on the stages of the policy cycle most amenable to injections of data-derived evidence. Höchtl et al. (2015) journey through the steps involved in policy making—e.g., agenda setting and discussion, policy formation and decision-making, implementation, etc.—providing reflections on the potential contribution of big data to each. They particularly highlight the possibility for real-time data processing to enable continuous evaluation throughout the process. Giest (2017) explores government use of a range of administrative and real-time data to design and customise policies. She highlights the value of these data to agenda setting and policy implementation. Malomo and Sena (2016) describe a case study of using integrated data in local government and highlight the benefits of big data for predicting need and effectively targeting services.

The studies usefully break down and compartmentalise the different functions of big data for policy making—options appraisal, predictive analysis, real-time evaluation etc. However, they tend to overplay the extent to which policymaking proceeds stepwise, through a series of linear stages, and understate the challenges associated with the straightforward inflow of any kind of information and evidence (Cairney & Heikkila, 2014).

Rather than seeing policymaking as a linear and rational process, we start from the premise that policymaking is the variable outcome of consensus, negotiation, contestation or co-option of ideas about what is to be done, by whom, how and for what purpose (Carmel & Papadopoulos, 2003). Ideas embodied in narratives of causation compete for the right to be accepted. Power and context influence the strength of the narrative to succeed (Jessop, 2009; Stone, 1989). Policymaking is a messy process in which conflicting ideas and policies are brought forward, debated, and implemented.

Scholarship is emerging on how data and data technologies fit into a narrative-conflict view of policymaking. Kettl (2016) emphasises that the nonlinear nature of policymaking problematises the assumption that data is used simply as evidence to make the best policy choice (see also Poel et al., 2018). They argue that good data analysis is useless without a good narrative. In contrast, Janssen and Helbig (in press) argue that data technologies have great potential to interrupt the status quo and revolutionise policymaking.

In summary then, the ideas that foreground government(s) enthusiasm for realising the potential of big, in-

tegrated forms of data have primarily focused attention on the potential of technical innovations. However, processes of data science in policy settings are embedded in dynamic, multifaceted, and deeply political contexts of problem definition, evidence interpretation, solution identification and decision-making. These settings materially affect the ways in which big data is called upon and able to impact decision-making. We engaged with local government activity around integrated data in order to consider how data informs policymaking processes: how the practices and preoccupations of the policy process define and shape the generation and use of data science; and how integrated data, as one form of evidence generation, shapes and redefines these policy practices.

3. Methods

This article presents a series of observations drawn from participatory action research within a set of local government data projects that ran at different times and for different durations between 2013 and 2018. Together, these projects set out to realise the benefits of integrated administrative and other data to policy development and practice at the local level, with the ultimate aim of establishing, testing and evaluating processes to change the culture of data use within and across public services.

Given the project aim, the approach was grounded in the principles of participatory action research (PAR) (Bergold & Thomas, 2012; Coghlan & Brydon-Miller, 2014) in a cycle of data collection and analysis, reflection and action, that emphasised equal collaboration between researchers and practitioners, trust and discretion in communication and the production of shared knowledge (Brydon-Miller, Greenwood, & Maguire, 2003). This approach was applied to four contemporary policy priorities for the statutory authorities involved (see Table 1). Working within the tenets of PAR, the research conducted within these four settings utilised data linked and anonymised by the statutory authorities prior to release for project purposes, and sought to contribute to both the development of data-informed policy and practice, and wider understanding of the contexts, processes and practices that realise the benefits of data for local government. In this article we present our observations from these projects.

The core project team that worked across all four policy priorities included three researchers from the University of Bath Institute for Policy Research (the authors) and three senior policy officials from two local statutory bodies within the South West of England. In the course of the projects the team engaged with service managers from relevant departments, and with other policymaking bodies and civil society organisations in the region. These included commissioning managers with responsibility for setting policy priorities; business analytics officers and managers from commissioned services and voluntary sector delivery bodies. The size and composition of the wider stakeholder group involved varied consider-

ably between projects, a point relevant to understanding variation in the conditions under which data-derived evidence can inform policy and practice and which we reflect on in the findings section below. The project activities were instantiated within a formal collaboration agreement between the three core institutions, which detailed data management and use protocols, and received ethical approval from the University of Bath.

Table 1 outlines the four settings for the research and the associated data sources used to inform decision-making.

In each case, the projects progressed through discussion with the project team and wider stakeholders to understand the policy issues and context; define policy questions of interest; identify and access potential sources of data; conduct and interpret quantitative analyses (e.g., propensity score matching, cluster analysis and predictive analysis). Insights from the analysis often raised additional questions, and policy questions were refined, and additional data and analysis sought accordingly. This process of making sense of the data and deciding next steps took place within regular fortnightly meetings of the project team at the University, as well as ad hoc meetings with other policy actors involved in each of the settings when each project was 'live'. In addition to the comprehensive notes taken of all of these meetings our reflections and observations draw on email exchanges, telephone conversations and the content of and comments on project documents (including, for example, project scoping documents and reports of the analyses).

4. Findings and Reflections

In drawing together the projects and seeking to explore the interactions between the policy context, policy questions and data integration practices, we present findings and reflections under three themes. Firstly, we consider the way in which the *relevance* of data is constituted in policy settings, as a function of its perceived value in answering policy questions. Secondly, we explore the conditions under which data applications to policy are *resisted*. Finally, we reveal significant aspects of the *relationships* between different interested parties where data and policymaking intersect.

4.1. Relevance of Data to Address Policy Questions

In using integrated data in local government settings, policy questions, not data, were the starting point for data projects. Whether the issue was financial hardship, designing health and wellbeing services or education service provision, it was the policy questions and context that defined the scope for data to inform decision-making. In this context, data did not "speak for themselves" (Anderson, 2008). Its potential utility to policymaking was realised where it was deemed able to be relevant to, and admitted (along with other evidence) as a

Table 1. Policy priorities, aims and data sources.

Policy Priorities	Policy Aims	Indicative Data Sources
Financial hardship	To understand the consequences of economic downturn and austerity for financial wellbeing.	*Individual-level:* • Time-series: social benefit claims, employment status, household composition, disability status • Demographic information *Aggregate-level (Lower Super Output Area):* • Debt (County Court Judgements) • Household composition • Social benefits claims • Tax credits • Income deprivation
Community health services	Review community health services for patients with a particular chronic condition to understand the efficacy of these services and the effect on health outcomes.	*Individual-level:* • Time-series: secondary care records, in-patient admissions, out-patient appointments, co-morbidities and clinical test results; • Attendance at community care services. • Demographic information *Aggregate-level (GP Surgery):* • Patient population; • Health checks. *Additional data collected:* • Patient illness perceptions and experience of services.
Wellbeing services	Review and redesign community wellbeing services.	*Individual-level:* • Wellbeing service administrative records: participant numbers, dates and service location; • Case notes; • Evaluation and outcomes. *Additional data collected:* • Interest in wellbeing services; • Various measures of personal wellbeing; • Demographic information; • Provider experiences of delivering wellbeing services.
Education services	Understand changed profiles of demand and redesign education services.	*Individual-level:* • Time-series: school and academies census; • Pupil demographic information; • Educational needs and status; • Free school meals eligibility.

response to a policy enquiry. In other words, the value, or otherwise, of data was constituted only in the context of critical examination of what data could represent and what it could say—given how it was generated, curated, processed and integrated—and only in relation to policy questions. Even where there were large volumes of exhaustive data, the application of that data to local policymaking was contingent on what data was considered able to illuminate about the perceived problem and what was allowed to be asked about it. The weight of data, typically associated with the big data phenomenon, did not unproblematically transfer into a weight of evidence (Schintler & Kulkarni, 2014).

Having said that, the projects do illustrate how a keen interest in the power of data, particularly the potential of combining multiple forms of disparate data, is reinvigorating and reshaping the demand for evidence in policymaking processes at the local level. Policy partners

were keen to identify and explore the benefits of the vast amounts of data routinely collected to inform service development, and were, in some cases, open to broadening the options for policy change in light of the subsequent insights.

There was sometimes an absence of data deemed sufficiently relevant to addressing particular policy questions. As an example we discuss the case of the review of local health services, which explored patient pathways and outcomes through services relating to a particular condition. In a routine appraisal of these services policy officials were interested in understanding barriers to and enablers of service take-up. They had a clear view about the nature of the policy problem: low levels of service take-up among certain patient groups in particular areas—and a set of questions predicated on assumptions about policy options for service improvement. However, project discussions with the research team led them

to broaden their enquiries. They commissioned a Rapid Evidence Assessment (REA) to extend their understanding of the factors influencing service uptake. A REA is an evidence synthesis that follows a systematic methodology but, in order to be rapid, is restricted in breadth, depth and comprehensiveness compared to a systematic review (Barends, Rousseau, & Briner, 2017). The REA raised explanations for low service take up and variation in service performance that were not previously part of the scope of the data project. This called into question the sufficiency of the data that had previously been designated as relevant to informing the policy question.

Policy makers became aware that data routinely collected and available on these services largely served to facilitate service administration and audit (e.g., by providing information on volume of provision, attendance and dates) rather than understanding reasons for low service take up review performance. They recognised gaps in the data relating to patient experience, as well as patient health management behaviours. In this case they decided to collect additional survey data. The survey drew together a number of existing validated scales (including the Illness Perception Questionnaire) and the sample size was all patients. The responses were combined at the level of the individual with existing administrative data to inform their decision-making.

In contrast, in other projects, the boundaries of the policy issue were broader and questions more loosely specified. For example, the enquiry into the consequences of economic downturn and austerity began with the broad aim to utilise linked data to identify changes in frequency and intensity of financial hardship at the local level. Equally, the review of education services began with a general aspiration to better understand changes in the profile of demand. In these cases, formulation of the policy questions and defining and deciding on the scope for data enquiries progressed through a series of incremental, iterative steps. Here, policymaking tended to be in response to emerging policy issues where there were numerous stakeholders advancing competing narratives about the nature of the problems and seeking to shape the range of acceptable policy responses. Thus unlike the healthcare case above, here the framing of the policy questions and legitimate solutions were contested. Despite policy officials' enthusiasm to realise the potential of integrated data, broadly defined questions raised challenges for identifying the types of data that could usefully provide answers. In the education service case, policy officials and service managers initially struggled to conceptualise how the various data on pupils and schools that they held could be exploited. The breadth of policy questions rendered the sources of relevant data that could address the questions as opaque.

In these cases, seeking to establish the existence and/or the relevance of data often involved conversations between the core project team and other data holders—often service managers in departments within the two local statutory authorities but outside the area of direct policy interest. This then involved a second-stage of iteration, to establish the validity of the data access request and legitimise the relevance of the data. In the health and wellbeing and the understanding financial hardship case studies, access to data held by other service providers was denied on the grounds that the resource cost of providing data was greater than the perceived benefit to policy. Combining data involves multiple sites where judgements are made about the relevance of data to policy questions that may not be owned or of interest to those that hold the data.

Issues of data relevance are also circumscribed by the divisions of local and national policy responsibilities. In the case of the data enquiry into the impact of economic downturn and austerity, the insights drawn from an analysis of combined datasets on levels of benefit claiming, employment status, county court judgements, household composition, physical health and other factors, showed particular groups of people (in work on low pay) as potentially more exposed to financial hardship. However, the ability of policy officials to action this insight was restricted, as it was deemed outside the scope of local policy. This case illustrated that insights from available and relevant data may not be actionable. This may be for a range of reasons—in this case, local government action was precluded by national government ownership of what transpired to be the issue where action was required.

4.2. Resistance to Data Use in Policy

The projects provided examples of ways in which the application of data to inform policy was challenged and resisted. For example, policy officials disputed or sought to discredit the legitimacy of data use where they had reservations about its quality. Sometimes claims about poor data quality were substantiated with reference to the purposes for which it had been generated: reservations were expressed around the notion that data collected for one reason should be repurposed for another. On other occasions resistance was focused on the way in which the dataset had been constructed where reservations focused on the validity of repurposing particular variables. Anticipation of public perceptions about the re-use of data also served to bolster concerns and augment resistance to data use.

In all of the projects, concern was raised about the potential impact on re-appropriation of the data of missing observations, human error and biases resulting from how they were collected, maintained and stored. In the financial hardship case study, policy officials resisted the inclusion of certain data fields on the grounds that the values they contained may be incorrect. For example, they questioned the quality of some demographic information in one data set where individual characteristics had not been crucial to determining service eligibility. Similarly, in the wellbeing services case, data related to the provision and uptake of these services (e.g., num-

bers of participants) were perceived to be more systematically collected—and thus more accurate—than evaluation data or data on participants' health outcomes. It was the evaluation and health outcome data, however, that was of greater value and significance in the re-appropriation of the data and the potential for linking with other data sets. Thus in both these examples, the extent to which data was considered suitable for reuse was related to the social context in which the data had originally been compiled: the likely motivations underlying the inclusion of particular variables and imputations about the care with which the data set had been constructed.

Further challenges to the validity of data applications for policy were raised in the education services case. Here the legitimacy of repurposing the data was less about the accuracy of the data and more about the validity of extrapolating from it. The example of data on eligibility for free school meals (FSM) illustrates this point. Even where data was perceived to be recorded correctly (i.e., all eligible registrations for FSM were input on data systems), policy officials highlighted that the introduction of universal infant Free School Meals in 2014 had significantly affected the numbers of parents registering their child's eligibility (Sellen & Huda, 2018). The perceived effect of this policy change was that FSM data had lost its value as an indicator of changed profiles of demand for education services.

In all of the cases, it was not that policy officials lacked curiosity and enthusiasm for harnessing the value of existing data. Indeed, aspirational ideas circulating within and beyond local government (e.g., Mayer-Schönberger & Cukier, 2013; Shakespeare, 2013) about the vast potential of big data permeated their thinking and motivated their efforts to realise the benefits for policymaking. However, the processes of data curation highlighted that the ability to be curious was tempered by the contexts in which datasets were compiled, structured and maintained in local government settings. For example, it was clear in the financial hardship case that a consequence of decisions to hold personal data on clients only for the time that they were service users was that datasets tended to over-represent continuous, and longer-term service users, thus obscuring patterns in short-term and cyclical service use.

To some extent the limitations inherent to data collection and management terms were perceived by policy officials to be a consequence of data protection compliance; specifically the requirements—under the Data Protection Act 1998 (Information Commissioner's Office, n.d.-a) and the (at the time forthcoming) General Data Protection Regulation and Data Protection Act 2018 (Information Commissioner's Office, n.d.-b)—to only collect and retain as much personal data as is necessary, and not to reuse data in ways incompatible with the original purpose. Where there were limits on data applications given the terms under which data had been generated, policy officials were reluctant to revisit consent

and tended to opt for the narrowest interpretation of their ability to generate or reuse data. This thus limits "extensibility" (Mayer-Schönberger & Cukier, 2013, p. 109), whereby the ability of data to have multiple uses is intentionally embedded in data collation protocols.

In addition, even where legal compliance was assured, policy officials were often juggling between two competing narratives about public perceptions of data use by local government. While they recognised a sense of public expectation that they would use available data 'smartly' to innovate and better target services, in practice they were also stifled by anticipation of public reservations about the acceptability of linked data. In other words, in their use of data policymakers recognised a distinction between what is legally defensible and what may be considered ethically permissible.

As a consequence, emerging awareness of data to answer policy questions did not unproblematically translate into availability of data. Policy makers' sensitivity to data quality and legitimacy, the legality of its use and the anticipated responses of the public could lead to data being rendered inadmissible in integrated data projects. Professional tacit knowledge was used to ground data, counteract its inaccuracies, navigate its ethical and legal implications and mitigate the likelihood of misreading the insights that it can yield. Data was only admissible where policy professionals could first fill in blanks and inaccuracies with their local knowledge of how things *actually* are.

4.3. Relationships with Data and Policy

This final section presents significant aspects of the relationships that effect the intersections between data and policymaking. We first observe that trust is vital to enable integrated data projects to have value in policy settings and then consider how the politics of policymaking impacted data sharing and the terms of engagement for different stakeholders.

Throughout the project collaboration, data was sourced and released in stages as trust in the partnership—between members of the core project team and the wider stakeholders—was built over time, ethical and legal boundaries established and the value of early analyses realised. For example, in the community health services case, establishing the policy-research relationship led to the project partners first seeing the potential value of conducting a RER, and then being confident to act on the relationship this showed between patient perception of illness and health management behaviours by collecting attitudinal data that could be linked with secondary health care records.

The data projects proceeded via an abductive approach—flip flopping between patterns emerging in the data and hypotheses, seeking additional insights and testing further hypotheses. For instance, in the example above, having refined the initial scope of the enquiry in the light of the RER, mini hypotheses to ex-

plain low service take up by certain patient groups were proposed, tested, discussed and revised in relation to the policy context. Across each of the projects, the rationale for additional data releases was grounded in the cementing of trust in the partnership and the realisation of benefits from the preceding stages. Thus, the value of the collaborative data enterprise was realised through processes that iteratively established confidence in the partnership.

Sometimes relationships between the project partners were more problematically embedded in the politics of data sharing; for example between levels and departments of government, between different public services, and between the policy partners and the research team. Some data—for example individual-level data on unemployment and take-up of employment services—were held nationally by the Department for Work and Pensions and unavailable to local policy officials on the grounds that it would breach their terms for information governance. Thus relevant data on variance in financial wellbeing was only available to the project in aggregate form.

On one occasion in the community health services project, difficulties in obtaining data from a service provider were attributed to the politics of the commissioner-provider relationship between the statutory authority and the provider. Given the nature of this relationship—and the unequal power relations within it—the senior policy officials within the core project team reflected that the other party may have been unwilling to share data for fear that the data would be misappropriated beyond the scope of the project and used to monitor their performance. This speaks to the significance of trust and transparency over purpose as well as methods in integrated data projects. Concern about the potential for data to surveil service performance was particularly apparent where ideas about policies—what they intend to achieve, for whom and how—were disputed. For example, in the wellbeing services project, service providers were unwilling to share data with service commissioners where they felt exposed when sharing data showing low volumes of activity without taking into account the quality of provision for vulnerable clients. A further variation on this theme was observed in the review of education services. Here data analysis was sought by service managers where it gave confidence to pursue preferred explanations for changed profiles of demand. Alternative explanations were undermined by questioning data accuracy or by citing particular aspects of policy context.

A final example from the financial hardship case, of the importance of trust was evident in a debate between one of the policy partners and a third sector organisation. The dispute centred on the scale of financial hardship in the local area and the nature of services required in response. Third sector providers made reference to a range of evidence to support their position. Significantly, the data held by these third sector providers was not made

available for integration as they claimed that its collection was conditional on particular sets of expectations for use. Their contention was that the data had been shared with them precisely because they were distinct from local government and a source of support for those wishing to raise grievances about local government. As a result they considered that sharing these data with local authorities would be a breach of trust. This provides a further illustration of how limits on linking data are not restricted to technical issues about the availability or format of data—rather they are shaped by relational considerations around trust and the politics of data and policymaking.

5. Discussion and Conclusions

The findings and reflections from our project to realise the benefits of data for policy have revealed particular sets of ideas about data (Markus & Topi, 2015). These concern the ways in which the relevance of data is socially constituted in policy settings and the conditions under which data applications to policymaking can be and are resisted, as well as the degree to which the relationships between stakeholders at the intersection of data and policy influence what data processes and insights can be considered. Overall, we highlight that variation in the degree to which integrated data and the techniques of data science are able to encroach on policy practice, is contingent on the ideas about and social context and processes of both data generation and policymaking.

The ambition to utilise the vast quantities of data that local government produces and can access is driven, at least in part, by the motivation to realise the aspirational claims made about big data for policymaking. However, the projects we draw on highlight the first-and-foremost requirement to be problem-oriented in big data applications to policy. Even where we observe the seeming ubiquity of data, there are still circumstances where we have data for which there aren't questions and questions for which we do not have data (boyd & Crawford, 2012; Kennedy et al., 2015); and it is questions and not data that drive policy calls on evidence.

In contrast to early definitions of big data that focused on the characteristics of data (volume, variety, velocity) with less reference to the purposes to which it could or should be put, we find that where integrated data is applied to policymaking its most defining quality is its ability to be big in value (Cowls & Schroeder, 2015; Organisation for Economic Co-operation and Development [OECD], 2013). In policy settings the value of data is allied to its ability to provide insight germane to live and pertinent policy and practice preoccupations. We find that the choice of what data to use or collect involves problem-based decisions on what would be indicative of the thing(s) we are trying to understand.

Given this grounding for the potential of data for policy, the social contexts and processes involved in data generation, maintenance and storage become of vital

importance. It is these contexts and processes that determine what data can, and what it cannot, represent and say. We have shown that administrative datasets tend to function primarily as a tool to audit public services; telling us how many services are delivered, for how many people and when. As such, their reuse value is limited where the aim of data applications to policy enquiry is the curious exploration of social phenomena, to understand what could work better, for whom and under what conditions.

The value of integrated data to policy challenges is further exacerbated when consideration is given to the errors and biases data contains as a consequence of how it is arrived at; what priorities are ascribed to its accuracy; and what legitimacy and legality it has when it is repurposed. The implication of these considerations is that the existence of large quantities of data is not an asset in itself to local policymaking. Its value can only be realised if and when the constraints of the social contexts and processes of its production can be mitigated. Even then, we have shown that the potential value of data is conditional on the political context in which policy is being made.

We have shown considerable differences in the contexts in which local policies are made. These contexts are not fixed and static, but highly variable, multifaceted and contingent on the historical trajectory of policymaking in the field. The context shapes ways of acknowledging problems and justifying the solutions to which policy is aimed.

Policymaking takes place on different timescales depending on the mode of policymaking. For instance, whether policymaking is happening as part of a routine programme of on-going review, or in response to an unanticipated shock—such as a public (media) outcry, a change in national or regional policy, or a change in social/economic circumstances—that disrupts routine policymaking processes and 'normal' policy timetables. At any given time, policy concerns can accelerate up through the rankings of priorities, or become suddenly subordinate to other more pressing preoccupations.

Big data analytics, with its focus on quick, novel and exploratory enquiry (Höchtl et al., 2015; Mayer-Schönberger & Cukier, 2013), could be seen to align well with extraordinary and fleet of foot policymaking; often seen as happening at a pace that traditional methods of information generation can't match (Whitty, 2015). However, such an assessment of the potential impact of big data-derived evidence underplays the complexity and politics of policymaking, particularly at points of disruption—for example times of economic downturn and austerity. In our experience, both times of routine policy appraisal and urgent reaction to policy crisis involve, first and foremost, the advancement and debate of *ideas about policy*, as well as related *ideas about data* (Markus & Topi, 2015) and what constitutes evidence.

The extent to which policy problems and potential options are tightly defined and agreed upon differs in different policy context. Ideas about policy, data and evi-

dence are contained within a political reality that shapes and delimits the boundaries of policy aims; the purpose to which it can be addressed, the extent to which ownership and responsibility over the domain is open or closed, and the degree of disagreement and dispute among stakeholders over the aims and purpose of policy. The nature of the policymaking context and the issues being explored affects what questions can legitimately be asked of big data and the ways in which the resultant insights are considered as admissible as evidence that can form the basis for decision-making. Issues vary in the degree to which they are contested, how urgent they are, how open, how risky, etc. As a consequence, we find that in practice highly contested local welfare policy has a qualitatively different profile of considerations shaping the 'pull' on data science than, for example, the temporarily more consensual context of local health service provision for patients with a particular chronic condition.

Thus in our exploration of how the practices of data intersect with the practices and preoccupations of policy, we find a more nuanced and politically contingent call on data than would be suggested by the rhetoric around the potential of data. Indeed, we suggest that rather than looking at data science as a technical aspect of government activity underpinned by expansive claims for the power of data, we should instead see data science as contingent on the ideas, realities and political contexts of government practice. Scholarship and practice around these topics must be alert to both the potential impact of data on policymaking but also the ways in which the practices of making policy condition the potential for data to be used.

Acknowledgments

Conflict of Interests

The authors declare no conflict of interests.

References

Anderson, C. (2008, June 23). The end of theory: Will the data deluge make the scientific method obsolete? *Wired*. Retrieved from https://www.wired.com/2008/06/pb-theory

Barends, E., Rousseau, D. M., & Briner, R. B. (Eds.). (2017). *CEBMa Guideline for rapid evidence assessments in management and organizations (Version 1.0)*. Amsterdam: Center for Evidence Based Management.

Bergold, J., & Thomas, S. (2012). Participatory research methods: A methodological approach in motion. *Forum: Qualitative Social Research*, *13*(1), 1–31. doi:10.17169/fqs-13.1.1801

boyd, D., & Crawford, K. (2012). Critical questions for big

data. *Information, Communication & Society*, *15*(5), 662–679. doi:10.1080/1369118X.2012.678878

Brydon-Miller, M., Greenwood, D., & Maguire, P. (2003). Why action research? *Action Research*, *1*(1), 9–28. doi:10.1177/14767503030011002

Cairney, P., & Heikkila, T. (2014). A comparison of theories of the policy process. In P. Sabatier & C. Weible (Eds.), *Theories of the policy process* (3rd ed.). Chicago, IL: Westview Press.

Carmel, E., & Papadopoulos, T. (2003). The new governance of social security in Britain. In J. Millar (Ed.), *Understanding social security* (pp. 31–52). Bristol: Policy Press.

Coghlan, D., & Brydon-Miller, M. (Eds.). (2014). *The SAGE encyclopedia of action research* (Vols. 1–2). London: SAGE Publications Ltd.

Connelly, R., Playford, C. J., Gayle, V., & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research*, *59*(C), 1–12. doi:10.1016/j.ssresearch.2016.04.015

Cowls, J., & Schroeder, R. (2015). Causation, correlation, and big data in social science research. *Policy and Internet*, *7*(4), 447–472. doi:10.1002/poi3.100

Giest, S. (2017). Big data for policymaking: Fad or fast-track? *Policy Sciences*, *50*(3), 367–382. doi:10.1007/s11077-017-9293-1

Höchtl, J., Parycek, P., & Schöllhammer, R. (2015). Big data in the policy cycle: Policy decision making in the digital era. *Journal of Organizational Computing and Electronic Commerce*, *26*(1/2), 147–169. doi:10.1080/10919392.2015.1125187

Iliadis, A., & Russo, F. (2016). Critical data studies: An introduction. *Big Data & Society*, *3*(2), 1–7. doi:10.1177/2053951716674238

Information Commissioner's Office. (n.d.-a). Guide to data protection. *Information Commissioner's Office*. Retrieved from https://ico.org.uk/for-organisations/guide-to-data-protection

Information Commissioner's Office. (n.d.-b). Guide to the general data protection regulation (GDPR). *Information Commissioner's Office*. Retrieved from https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr

Janssen, M., & Helbig, N. (in press). Innovating and changing the policy-cycle: Policy-makers be prepared! *Government Information Quarterly*. doi:10.1016/j.giq.2015.11.009

Janssen, M., Konopnicki, D., Snowdon, J. L., & Adegboyega, A. (2017). Driving public sector innovation using big and open linked data (BOLD). *Information Systems Frontiers*, *19*(2), 189–195. doi:10.1007/s10796-017-9746-2

Jessop, B. (2009). Cultural political economy and critical policy studies. *Critical Policy Studies*, *3*(3), 336–356. doi:10.1080/19460171003619741

Johnson, J. A. (2015). What is the data in big data? *Discover Society*. Retrieved from http://discoversociety.org/2015/08/03/on-the-frontline-what-is-the-data-in-big-data

Kennedy, H., Moss, G., Birchall, C., & Moshonas, S. (2015). Balancing the potential and problems of digital methods through action research: Methodological reflections. *Information, Communication & Society*, *18*(2), 172–186. doi:10.1080/1369118X.2014.946434

Kettl, D. F. (2016). Making data speak: Lessons for using numbers for solving public policy puzzles. *Governance*, *29*(4), 573–579. doi:10.1111/gove.12211

Kitchin, R. (2014a). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1), 1–12. doi:10.1177/2053951714528481

Kitchin, R. (2014b). *The data revolution: Big data, open data, data infrastructures & their consequences*. London: Sage Publications Ltd.

Kitchin, R., & McArdle, G. (2016). What makes big data, big data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, *3*(1), 1–10. doi:10.1177/2053951716631130

Malomo, F., & Sena, V. (2016). Data intelligence for local government? Assessing the benefits and barriers to use of big data in the public sector. *Policy and Internet*, *9*(1), 7–27. doi:10.1002/poi3.141

Markus, M. L., & Topi, H. (2015). *Big data, big decisions for science, society, and business*. Waltham: Bentley University. Retrieved from https://www.bentley.edu/files/2015/10/08/BigDataWorkshopFinalReport.pdf

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data. A revolution that will transform how we live, work and think*. London: John Murray.

Organisation for Economic Co-operation and Development. (2013). *Exploring data-driven innovation as a new source of growth: Mapping the policy issues raised by 'big data'* (OECD Digital Economy Papers, No. 222). Paris: OECD Publishing. doi:10.1787/5k47zw3fcp43-en

Poel, M., Meyer, E. T., & Schroeder, R. (2018). Big data for policymaking: Great expectations, but with limited progress? *Policy and Internet*, *10*(3), 347–367. doi:10.1002/poi3.176

Schintler, L. A., & Kulkarni, R. (2014). Big data for policy analysis: The good, the bad, and the ugly. *Review of Policy Research*, *31*(4), 343–348. doi:10.1111/ropr.12079

Sellen, P., & Huda, N. (2018). Evaluation of universal infant free school meals. *Education Policy Institute*. Retrieved from https://epi.org.uk/publications-and-research/evaluation-universal-infant-free-school-meals

Shakespeare, S. (2013). *Shakespeare review: An independent review of public sector information*. London: Department for Business, Innovation and Skills. Retrieved from https://www.gov.uk/government/publications/shakespeare-review-of-public-sector-information

Stone, D. A. (1989). Causal stories and the formation

of policy agendas. *Political Science Quarterly*, *104*(2), 281–300. doi:10.2307/2151585

United Kingdom Parliament. (2015). Making big data available: Key issues for the 2015 Parliament. *United Kingdom Parliament*. Retrieved from https://www.parliament.uk/business/publications/research/key-issues-parliament-2015/technology/big-data

Uprichard, E. (2013). Big data, little questions. *Discover Society*. Retrieved from https://discoversociety.org/2013/10/01/focus-big-data-little-questions

Whitty, C. (2015). What makes an academic paper useful for health policy? *BMC Medicine*, *13*(301), 1–5. doi:10.1186/s12916-015-0544-8

About the Authors

Hannah Durrant leads the research programme of the University of Bath Institute for Policy Research. Her research interests are the governance of economic and social policy reform and the role of data, evidence and expertise in policymaking. She is currently working on several projects related to big data in government.

Julie Barnett is a Professor of Health Psychology at University of Bath. Her research interests and projects span public and stakeholder understandings of risk, risk communication, management of food allergy and intolerance, loneliness and the way in which new forms of data are used as evidence in policy making.

Emily Rempel is a Researcher at the University of Bath in the Department of Psychology and the Institute for Policy Research. Her work focuses on the intersections between publics, policy, and data technology with an emphasis on feminist perspectives. Other research interests include health and well-being, epidemiology, and population mental health.