THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Datafication of Government:

Mapping Data-driven practices in Social Service Delivery

Asha Susan Titus

This thesis was originally submitted to the London School of Economics and Political Science for the degree of Doctor of Philosophy (PhD), April 2023

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it). The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent. I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

Word count: 76110

Abstract

The introduction of processes of datafication is profoundly changing the logic of care and welfare provision within the social services sector in the United Kingdom. This thesis presents findings from one of the first detailed empirical studies of the incursion of dataism (Van Dijck 2014) into sensitive policy fields like child protection, welfare and social care administration. I discuss the consequences of using measurable types (Cheney-Lippold 2017) to pinpoint problematic behavioural characteristics that might need interventions in the future. In doing so I contribute to discussions about how the digitalisation of government (Dunleavy and Margetts 2015), as a long term structural macro process, has produced changes in the underlying rationalities of government. This is accomplished by locating datafication as a historical process that follows digitalisation (Mayer-Schönberger and Cukier 2013, Couldry and Hepp 2016) and mapping the discourses and practices of data driven technology in government.

Drawing on in-depth case studies developed through fieldwork (conducted between 2020 to 2021) and engagement with key public sector departments, I demonstrate how welfare provision is being reconstituted through data. Risk based prioritisation (Yeung 2018, Yeung and Lodge 2019) of casework through data signatures represents a significant departure from earlier modes of working with vulnerable benefit claimants and defining what 'high risk' looks like. The majority of the published investigations of algorithmic processes in statutory safeguarding tends to focus on predictive risk assessment in the US context or on routine Child Protection and Child in Need functions of local government. This thesis offers a new empirical site and a few distinctively under-researched modelling processes to explore how the imperatives of datafication interact with existing organisational cultures and the mundane administrative interests of those who seek to optimise bureaucratic workflows.

I also document the justificatory discourses that accompany the introduction of data driven pilots and elicit the views and reflections of the human frontline worker on whether their role has subsequently shrunk. The main contribution of this thesis lies in the conceptually innovative manner in which it brings together the problematisation of 'data' from Critical Data Studies, classification theory and research from Social Policy.

Acknowledgements

Conducting research from the heart of central London at the London School of Economics and Political Science presented various challenges. This was especially the case as I worked on this research and manuscript through a global pandemic. I would like to thank my supervisors Dr Damian Tambini and Prof Nick Couldry for their guidance through this process. Damian has been a very good discussant when the arguments of the theoretical work took unusual and unexpected turns; and provided much needed moral support. Nick shaped the ideas around datafication and engaged deeply with the intellectual endeavour. Thanks is mainly due to the examiners of this project, Prof Eileen Munro and Dr Morgan Currie, for their generous feedback and insightful responses to the presentation of this work. Prof Ellen Helsper who took over as chair of my committee towards the last part of the journey asked all the right questions at the right time to help me complete this monograph.

I want to especially thank the multidisciplinary Open Science researchers at the Turing Way within the Tools, Practices and Systems hub of the Alan Turing Institute who I had the privilege of collaborating with during the final two years of my PhD. Through a book chapter writing grant, I contributed to an ethical guide aimed at a Data Science practitioner audience; but developed a sense of belonging which is rare to come by in a fast-moving space that navigates disciplinary boundaries. During a significant impasse in this doctoral journey, I found their supportive, inclusive and culture shifting approach to building communities of practice around the development of data science incredibly inspiring. Despite the criticisms directed at the biases perpetuated by the callous deployment of Data Science, their unwavering commitment to bringing together Life Scientists, Legal Sociologists and Data Ethics scholars provided me with a vibrant intellectual community and helped me gain momentum. In this regard, I commend Dr Kirstie Whitaker and Dr Malvika Sharan for coordinating and building this space.

Informal CiA writing groups established through the Social Science Faculty network at the University of Oxford provided encouragement that helped me tide over some rough patches during the course of this research. I feel gratitude to those I was introduced to, especially Sir Andrew Dilnot and Dr Grant Blank (at the Oxford Internet Institute), as they summarised the national social care policy context and provided me with longitudinal quantitative datasets on the demographics of who uses e-government with a representative sample of the United Kingdom.

One of the crucial components for the success of this project was negotiating fieldwork access. It is hard to sufficiently thank and acknowledge everyone who facilitated introductions as there were so many. Dr Juliette Malley at the Care Policy and Evaluation Centre (LSE), colleagues from LARIA (Local Area Research + Intelligence Association), Rebekah Wilson at the Local Government Association, Tom Rintoul at Social Finance, Paul Waller (University of Leicester) and Baroness Philippa Stroud generously gave of their time and put me in contact with key stakeholders.

Finally, my utmost gratitude to the participants of this study for being willing to engage and reflect on their day-to-day work practices as civil servants. It is not possible to express my thanks to various expert interviewees and the many respondents who contributed to the background policy context mapping work without breaching confidentiality agreements. Based on our interactions, I remain optimistic that when the findings of this project are made available to the 'data workers' closest to these processes of datafication, they will want to do everything they can to incorporate the critique; and effect lasting change in order to help the vulnerable welfare recipients they serve. I add the caveat that this would only be possible if there is the political will to do so at senior levels of the bureaucracy within the government agency and Whitehall leadership apparatus.

I would also like to acknowledge the invaluable support of Prof Rita Astuti and the LSE PhD Academy team for supporting this research with various kinds of institutional and financial assistance. I conducted ethnographic fieldwork amidst a global pandemic while living in London through 18 months of intermittent COVID lockdowns. As an Australian student whose government had imposed severe border restrictions throughout 2020 to 2021 this posed a range of challenges. Their financial support and flexibility enabled the successful completion of this project. For all comments and feedback received in conferences where I have presented earlier version of the findings, many thanks is owed to my academic interlocutors.

Doing interdisciplinary, policy relevant research is not easy. It would be dishonest for me say that the Department of Media and Communications at the LSE made a good intellectual home for this project. I found the research culture of my department hypercompetitive, cold, harsh, needlessly critical, riddled with unconscious biases and counter-productive to my progress. It never ceases to amaze how senior critical scholars who have spent their entire careers studying power, seem to be oblivious to its operation in their own work practices and relationships with junior scholars and PhD students. I was however delighted to have found the support and

friendship of Rob Sharp, Vanessa Ciccone and Ludmila Lupinacci from amongst my cohort. Prof Sonia Livingstone helped blaze a pathway forward and took me to concerts when I needed cheering up! I am surprised by my resilience and determination; but remain convinced that so much more needs to be done to make academia less hostile to women of colour, those from racialised minority and other non-traditional intersectional backgrounds. It remains my hope that at least in the publications and policy recommendation white papers that come from this doctoral work, you would cite me, respectfully handle my intellectual contributions and not undermine me by merely utilising it as part of a literature review without proper attribution and credit.

This thesis has benefitted from the support of numerous colleagues and friends over the years. Being part of the Warden's team at the LSE providing care and support to graduate students through the latter half of my PhD tenure, there are many who covered my shifts before important deadlines so I can carve out focused blocks of time for writing. Max Goehmann, Sarah Lee, Glen Gostlow, Yohan Iddawela, Prakash Pandya and Rachel Zhou stand out in this regard. Much needed inspiration and sustenance came from the groundbreaking work environment and goal of the Ada Lovelace Institute to make algorithms accountable to those who are most negatively impacted by them.

Overall, the guiding spirit for my research investigation was my hope that vulnerable individuals from low-income backgrounds, with entrenched intergenerational and complex experiences of poverty wouldn't be reduced to mere metrics. The disruptions caused by disability, in-care experience, sexual abuse and domestic violence can be profound across the lifecycle. In mapping data driven practices in statutory safeguarding, welfare provision and other administrative scenarios within central and local government, I sought to ensure that the data infrastructures that policy makers build and use would reduce harm to vulnerable populations; rather than reify, entrench and compound the conditions of their disadvantage. It is to this *objectif générel* that this research and thesis documentation contributes.

Table of Contents

Summary of Theoretical and Analytical contributions - A differentiated understanding of Data

Chapter 1

1.1 Introduction

Datafication refers to the growing impulse to convert aspects of the world that have been outside the purview of processes of quantification into digital data. From footsteps, vital signs and health metrics to web browsing patterns, interactions and subjective experiences, almost all life processes can now be rendered into data form. When converted into indexical, machine readable datasets, these become amenable to search and sort functions, aggregation and large scale computer processing. Within this data landscape, many have argued that 'new kinds of value' can be extracted from converting information generated for one purpose to another with the ultimate aim of mapping the world in a 'quantifiable, analysable way' (Mayer-Schoenberger and Cukier 2013:97). The claim is that aside from the seemingly technical enterprise of converting a phenomenon into a 'quantified format so that it can be tabulated and analysed', the move to Big Data is a 'continuation of humankind's ancient quest to measure, record, and analyse the world' (2013:78) with datafication being a 'great infrastructure project' that rivals the epochal shift signified by Enlightenment's *Encyclopédie*.

The normative values underpinning this universalising vision are clear. Unwavering optimism attached to such a data deluge led to claims about a new scientific paradigm (Anderson 2008) where it would be possible to do away with theory and rely on data to directly measure and understand any given phenomenon. In addition to those who took these claims about big data seriously, journalists, business strategists and technologists produced breathless accounts of the possibilities opened up by trend mining approaches and the computational tools that capture exhaustive datasets. The assumption is that the sheer size of this kind of data (where n = all rather than when a sample from a population) compensates for their messiness and minor variations are ironed out given the volume one is dealing with. This forms the basis for a shift from statistical inference based on a hypothesis driven approach using identified variables of interest to a trend mining approach based on correlations. The salient examples include identifying flu trends using Google search keyword datasets and providing personalised Amazon book recommendations after capturing browsing patterns, dwell time and metadata from one's myriad online activities and tallying it against user segmentation and targeted demographic profiles to compute consumer preferences (Cohen 2015, Barocas and Selbst 2016).

In much of the hyperbole surrounding the virtues of big data, ‘data’ always stands for something larger than itself, and its meaning is often mobilised and used as a symbolic and cultural resource to advocate for technological changes and present them as desirable. Van Dijck (2014) sees a kind of epistemological conceit in this faith in the objective nature of data as representing reality and offering a new gold standard of knowledge about human behaviour by calling it ‘dataism’. Critical approaches have pointed out how calls to adopt ‘data driven’ technologies employ slippages that blur the distinction between data collection and interpretation (Gitelman 2013) by implying that data points or the data wrangling process speaks for itself and produces self-evident, superior insights. In actuality, access to information we couldn’t collect before leading to large scale datasets, doesn’t necessarily speak for itself. In other words, patterns in digital trace data don’t seamlessly reveal human behaviour, motivations, sociality, relationships, sentiments and preferences – rather, narrative and ideological work goes into firstly, establishing the desirability of such quantification; and secondly, in the process of sense making through which meaningful patterns are identified as accurately ‘standing in’ for human behaviour.

Throughout 2016 and 2017, while working for government within the Australian context, I noticed the growing appetite for innovative data technologies and the euphoric use of ‘data-driven’ as a buzz word that kept appearing in project proposals. Appeals to ‘data’ were always made to strategically justify a policy shift and shore up the position and legitimacy of the project proponents. Data driven technologies were increasingly being positioned as the new frontier of public sector modernisation. There was a great deal of expectation attached to the promises of these data driven systems to augment decision making by providing new insights, real time feedback, user segmentation and targeting. Starting with the claim that there is value to be extracted from the ‘raw resource’ of administrative data trapped in outdated government IT systems that are waiting to be processed, repurposed and linked, I saw the emergence of the logic of dataism everywhere.

Data Insights Units were propping up especially close to Treasury and Planning departments and I noticed that these ‘data driven’ pilots were occupying the same space that Behavioural Nudge projects (John 2018) did previously; with young ambitious civil servants positioning themselves within these units in order to fast track their career. Data was indeed being used as a symbolic resource standing in for something larger than itself; with these units retaining the disruptive ‘fail fast, fail often’ experimental culture which was purposefully at odds with the values and ethos of the wider public sector. New kinds of evidence was being used for policy without questions being raised about their fundamental appropriateness.

For example, in order to gauge the planning needs of a fast-growing metropolitan region, one of these units built a data model that ingested Uber movements, rental and utility bills, commuter movements logged through the cards used to access the Underground (Subway) rapid transit grid in order to figure out who lives where, and with whom in near real time. Different neighbourhoods (some overpopulated) were not being serviced well enough by current public services and the aim was to use this real time information to make resource allocations based on a model that assessed whether suburbs had enough schools, hospitals and other such services. The aspiration was to get movement patterns down to 30-minute intervals as they claimed that census information was dated and not useful for planning purposes. The privacy implications of pinning down the whereabouts of residents of that city down to 30-minute intervals were brushed aside by reference to the innovative potential of such a data driven exercise. Rationales included the expectation that while this was just an initial pilot, there would be techniques that would develop later on that can anonymise citizen names and configure a way to only extract metadata from private records such as rental contracts and utility bills. Using material that was never previously thought of as evidence for policy, deeply personal insights about who lives where and with whom in near real time was compiled. Combining information about households from disparate datasets might not appear to be a direct intervention into the life of an individual but it is far from innocuous. This is especially the case here as it is impossible to know for certain the data quality of private textual records such as rental documents that are created for a specific purpose, audience and documentary function; which is then ingested by a large inscrutable model that combines it with other disparate pieces of information to provide demographic profiles for planning purposes.

In what was dubbed by the media as the ‘Robodebt’ debacle, Australia’s online compliance initiative (OCI) was a large scale data linking project that crystallised my discomfort with these technologies that initiate fundamental changes in administrative procedure; yet are framed as an inherently neutral, technical innovation. This was a debt recovery system that used fuzzy matching to link fortnightly income reported to the welfare agency Centrelink to yearly tax assessment information held by the Australian Taxation Office (for the 2010–17 financial years) with the names and often outdated addresses held on the system of vulnerable people on disability and unemployment benefits[1]. Based on the averaging algorithm that estimated that income was earned evenly across each financial year rather than on verifiable actual fortnightly earnings, when there were discrepancies in income reported, automated debt notices were sent out. The burden of proof shifted on to the welfare recipient to dispute the debt call by uploading old payslips through an online portal.

Frontline bureaucrats were instructed not to correct the many invalid debt calls and protestations that the benefit recipient had since moved or did not retain old payslips were not considered. This led to a situation where there was widespread distress and suicides and it became apparent that the might of the government was being used to bully low-income families. Despite Senate Inquiries and Administrative Appeals Tribunals[2], it was hard to pinpoint who to hold accountable even as it became obvious that the data science proposals that were initially presented to senior executive staff in the civil services were not properly understood by them. The belief in the veracity of the modelling scaled up a punitive technology leading this to become a cautionary tale[3] for how not to do administrative data innovation while dealing with vulnerable populations.

---

[1] The Department of Human Services (DHS) sent approximately 80 million identities to the Australian Taxation Office and received approximately six million matches back. This is because the DHS data file includes identity information such as names, date of birth and historical addresses and does not include the customer’s tax file number (TFN) because use of the TFN is restricted to data matching under the Data Matching (Assistance and Tax) Act 1990 (the Data Matching Act). ATO used tax file numbers to extract information it holds about that individual (for example, name, address and date of birth) and then retrieved the income reported in a manner that flouted administrative law. This generated approximately 20,000 income data-match discrepancy letters per week and the Department of Human Services undertook around 783,000 assessments in the 2016-2017 year. This stands in contrast to the 20,000 compliance interventions under the manual process. Recipients were asked whether those details were correct and the only way to dispute the debt call was by uploading old payslips as proof. Many felt unable to contest these debt calls or pay the exorbitant amount the system said they owed that had accrued over many years. For further detail on how legal aid and advocacy groups contested this leading to a Robodebt Royal Commission several years later, see my case study for the Turing Way (2020) or Carney (2018).

[2] Documentation from these enquiries reveal harrowing tales of repeated debt calls being sent to the families of those who committed suicide with remedial avenues closing up. The Royal Commission heard several cases where mothers of adult sons on disability benefits who had committed suicide found it particularly triggering when these letters with automated debt notices were then sent to their addresses following the lack of response from the intended, now deceased, recipient.

[3] With the proliferation of new data intensive methods that have been introduced into public sector organisations everywhere, this is now the case even beyond the Australian context.

When mainstreamed into day-to-day government functions and decisions are centred around data based representations of the world, this enables different kinds of interventions into civic life that is worth examining in careful detail. This research project has its origins in this personal observation (and frustration) that novel computational metrics are now being incorporated into government with welfare systems needing to embed their duty of care, the precautionary principle, and a 'do no harm' ethos into complex data infrastructures.

Technological change in bureaucratic systems is an under-studied area which has occupied a somewhat peripheral position within public administration and public management research despite producing long term structural changes in the underlying rationalities of government (Pollitt 2010, Dunleavy and Margetts 2015). However misleading or hyperbolic the narratives that extol the virtues of data, governments are increasingly buying into the promises of these technologies. The changes activated when processes of bureaucratic decision making are predicated on continuous data aggregation have a very real impact on the lives of the people who interact with the department or public agency.

Collecting, combining and processing information about individuals and organisations has always been a key function of government. What has changed is the scale and complexity of data operations, the availability of new sources of data and the reliance on large-scale aggregation and data association. In addition to public datasets which include descriptive statistics about populations, social indexes and economic indicators, it is now possible to collate a ‘data trail’ or a comprehensive electronic record from administrative transactions of one’s past decisions and interactions with government at an individual level. Linking this to traditional databases creates an information infrastructure that marks a change in kind and not just in degree. Drawing on other streams of digital data from sensors, mobiles[4], imaging technology and other trace data emitted as a by-product of the mundane use of technological devices adds another layer of complexity. With the radical expansion and embedding of digital technologies into every sphere of life, what is available to be measured is ever increasing. For example, fine grained ‘behavioural metadata’ can be mined from digital traces and aggregated risk profiles produced on the basis of that represent a drastically new source of evidence for policy making.

---

[4] Social media content has been streamed through mash ups that use geographic data to help manage traffic congestion and map commuter patterns (McNeill 2017)

It is important to foreground how in terms of its materiality, the characteristics of data elements we are talking about when we refer to processes of datafication, are distinctly different from previous modes of classifying, categorising and ordering information. Until the end of the twentieth century, digital records and systems looked almost exactly like their analogue, paper versions (Robertson and Travaglia 2015). We are living through a profound transformation where processes of governance are being reconstituted through data intensive methods that order and classify populations in new ways. It is therefore necessary to historically situate this moment of datafication. Kitchin and McArdle (2016) draw our attention to the volume, variety, velocity and exhaustibility of new forms of data and the socio-material arrangements and infrastructures of their production. Below are the key traits that they identify based on a comprehensive review of published work (2016:1):

- ‘volume (consisting of enormous quantities of data);
- velocity (created in real-time) and;
- variety (being structured, semi-structured and unstructured);
- exhaustivity (an entire system is captured, n = all, rather than being sampled) (Mayer-Schönberger and Cukier 2013);
- fine-grained (in resolution) and uniquely indexical (in identification) (Dodge and Kitchin 2005);
- relationality (containing common fields that enable the conjoining of different datasets) (Boyd and Crawford 2012);
- extensionality (can add/change new fields easily) and scaleability (can expand in size rapidly) (Marz and Warren 2015);
- veracity (the data can be messy, noisy and contain uncertainty and error) (Marr 2014);
- variability (data whose meaning can be constantly shifting in relation to the context in which they are generated) (McNulty 2014)’.

Couldry and Hepp (2016) provide a key theoretical intervention by locating digitisation as a neat precursor to datafication but also distinct from it. After mapping the progressive deepening of the reliance of all social processes on infrastructures of communication over time, they read history as waves of mediatisation which include Mechanisation, Electrification, Digitisation and Datafication. They posit that we are currently living in an emerging form of social reality characterised by deep complexity, moulded by media forms and where ‘data-processing becomes entangled in the emotions of everyday life’ (2016:139). Within the current phase, almost every interaction generates continuous, automatically produced data that is then handled by ‘processes of aggregation and algorithmic calculation’ (2016:125).

Terms like ‘third order technology’ (Floridi 2014) have been used to describe the new environments that datafication creates. Data lends itself to dematerialisation whereby information can be stripped of its physical context, moved, unbundled and re-bundled in a different context. In contrast to analogue modes of record keeping, it flows and circulates differently and there is a burgeoning literature focussed on delineating its key traits (Florescu et al. 2014, Lupton 2015, Kitchin and McArdle 2016). In the empirical chapters that follow, I develop a detailed account of how datafied models are being introduced at different levels of government. It is these novel properties of datasets - whereby they can be taken from one context, stripped of its physical contextual attributes, repurposed, and recombined with datasets from another context - that makes these models possible. In fact, combining levels of data about an individual by extracting fragments of data on the person held in different databases through complex processing and layers of abstraction are what generates the risk scores in Chapter 3. It also underlies the vision of ‘joining up’ data points held by different departments to deliver ‘slick services’ personalised around life events (Chapter 5). Chapter 7 also returns to this point when it addresses how large numbers of welfare claimants are partitioned and routed into a pipeline of automated A/B tests with relative ease based on data from the digital service logs that are amenable to filtering, sorting and flowing through these pipelines in an unprecedented manner.

Without paying heed to the new materiality of data produced under this phase of datafication through cheap memory, powerful processors, and advanced computational capabilities, it is hard to grasp what has changed from previous analogue modes of record keeping. Assumptions about what is real and what matters change with the proliferation and ascendance of new data technologies such as predictive analytics, algorithmic procedures and machine learning models that quantify and format the world in biased or unexpected ways. Drawing on this theoretical discussion of datafication and dataism, the next section builds a conceptual framework to locate the shift underway when foundational processes of government are predicated upon data intensive methods that group, divide and classify people.

2 Conceptual Framework: Categories, Administrative practices, and the expansion of processes of Datafication

> People come to street-level bureaucracies as unique individuals with different life experiences, personalities, and current circumstances. In their encounter with bureaucracy they are transformed into clients, identifiably located in a very small number of categories, treated as if, and treating themselves as if, they fit standardized definitions of units consigned to specific bureaucratic slots. (Lipsky 1980: 59)

In the above influential public administration book published almost 40 years ago, Lipsky (1980) sets out how a key tension in frontline welfare service work revolves around standardisation. This refers to the need to be responsive to individuals with varying backgrounds, needs, and service claims while being able to demonstrate that they are all being treated equally according to the rules and resource limits of the organisation. He postulates that the only way in which this can be achieved is through routinisation or the development of organisational routines and classificatory practices that convert events, people and things into entities that frontline bureaucracies can recognise and process. The category in this case becomes an 'institutional machinery' used to lump similar people (and problems) together and works as a 'kind of stabilising standard' (Douglas 1986:59). It is used to encode meanings, organise information and coordinate practices. In this sense, rules that are composed of deliberate categories that divide and order people by their identity, behaviour and situations are essential to understanding this level of governance and policy implementation (Hjörne and Säljö 2004, Stone 2005).

In addition to providing the means for institutional standardisation, state defined categories are routinely used to segment populations in order to target service delivery. Classification schemes based on seemingly value-neutral, empirically observable attributes such as age[5] and gender are almost always used as eligibility criteria for accessing a range of public services. Categories are important in the construction of 'at risk' groups as objects of preventive intervention (and in models of normalcy against which risky behaviour is defined). Target populations are implicit in the design of policy interventions and social categories tacitly inform 'whom the welfare state should worry about' (Østergaard Møller and Sommer Harrits 2013). The set of meanings and labels of deviant behaviour associated with these categories

---

[5] Like pension

are as important as observable criteria for membership in problematic groupings like ‘troubled families’, ‘children in need of additional support or early attention’, being from a ‘weak parenting background’ or ‘welfare cheats’.

Furthermore, subcategories are regularly redefined and deployed with consequence for how one should be treated and in which way by public policy. This means that demarcations and the boundaries between categories have to be actively maintained, negotiated and policed. For example, the lines that divide the ‘deserving poor’ from the ‘undeserving poor’ for the distribution of benefits and burdens are often discursively constructed. Therefore, policy making and implementation has routinely used nuanced classification practices to flatten out variation, unmanageable specificities and contextual details to deliberately sort, order and allocate people into artificial classificatory types.

We see how such ‘category making’ especially within an institutional setting requires work and does not reflect natural divisions that occur in the world; but involves a series of interpretative choices that produce order through differentiations and hierarchical[6] relationships. This process of demarcating, segmenting, and establishing similarity between individuals with different life experiences so that they are identifiable by street level bureaucracies is not entirely new; and has been in place since the pre-datafication era. Now, however, new streams of data are being used to group and divide people within social service delivery with classificatory tasks taking on a different order of importance.

---

[6] Classificatory systems have a clear defining point of view. As highlighting some similarities and occluding other features of difference forms a core part of classification, then it is important to think about who gets to define relevant categories and whose point of view prevails within these taxonomic systems. Which characteristics are being used to formalise the category and which characteristics are being left out? Within an institutional setting, the point of view that is adopted could just be the collective opinion of a majority, or those in power or what is widely perceived as habitual and legitimate behaviour. Empirical studies of categorisation practices within professional welfare work show how tacit preferences manifest themselves in allocative choices and comparisons. Stone (2002:53) delineates the role of ‘particular group comparisons’ in how judgements are made and rationales given for who is entitled to welfare support even after being assigned to clear-cut, ‘scientific’ categories. In what she describes as ‘category work’, she shows how decisions about whether a child suffering from ADHD is classified as disabled or as a product of bad parenting result in two very different sets of rights with regard to welfare support. This ambiguity is managed by the discretionary practices of frontline bureaucrats who place the specific case into an appropriate category by comparison to children with learning disabilities or to misbehaving children in general.

With the increasing reliance on computation and the growing incursion of processes of datafication into public administration how are datafied category parameters being set? More specifically within the context of social services and welfare provision, how are allocative decisions made about whether a person fits an old, existing administrative category when the evaluative criteria is now based on digital patterns and information signals derived from parsing through enormous quantities of information in real time? The legitimation of new data sources as a basis for policy and qualitative changes in how records can be aggregated, linked and analysed can lead to new ways of fitting cases into categories, filing and recalling information that can transform these core tasks of public administration. Digital data when used in recordkeeping lends itself to distinctive mechanisms of aggregation, filtering, association and sorting and represents a change in kind rather than in degree. When sophisticated, automated, fine grained segmenting techniques are applied to administrative databases what kinds of new category structures emerge and with what consequence?

John Cheney-Lippold in his book 'We are Data' (2017) introduces the concept of 'measurable type' to refer to categories that one is assigned to within such a data infrastructure based on 'algorithmic fit'.

> *A measurable type is a data template, a nexus of different datafied elements that construct a new, transcoded interpretation of the world.*

The classificatory process in this case is based almost exclusively on what is available to measure in a digital format. Seeing the world (and populations to manage) through these data based categories involves discarding contextual details from the close analysis of unit cases in the social world and relying on abstract, inferred characteristics.

I argue that processes of classification are intensified, automated and amplified with datafication and focussing on the following themes will allow me to deepen and develop an account of the growth of datafied administrative practices in welfare bureaucracies. Drawing on this notion of a 'measurable type', and situated at the intersection of classification theory and Critical Data Studies, I identify the following 4 analytical themes as the framework for this study (sections 2.1 to 2.4). As such, datafication is an emerging and shifting phenomenon. In order to pin it down and study at depth this juncture at which processes of datafication are gaining momentum and making inroads into even the most sensitive social policy fields such as welfare provision and child protection, a conceptual framework is required that identifies what it looks like and draws it into sharper focus.

A measurable type is a data based compilation of the category it supposedly represents. The complexity of individual histories are translated into a neat digital format when one is assigned membership in a 'gender' category based on web surfing patterns or allocated into an 'at risk' category compiled from digitally observed problematic behaviour. Extending Bowker and Star's scholarship (Bowker and Star 2000), Cheney Lippold (2017) argues that categories we are assigned to are often different to what we use for ourselves when being identified based on data. Therefore, there is a very clear distinction between 'female' as an algorithmically produced category and female as a self-described identity based on someone's lived experience and formed through conscious interaction with others in real life. Cheney-Lippold shows how browsing behaviour and a constant stream of real time web use data is matched against existing models to make math based identifications that have very little to do with actual gender. He differentiates between the measurable type of 'gender' and the non datafied identity of gender by the presence or absence of single quotation marks.

Data based assessments of someone's gender is assessed (often by marketers) based on database queries and code; and when the browsing patterns and keywords change, the dynamic and modulatory nature of the definition is such that 'the gender of the same user might change from male to female' (Cheney-Lippold 2011). Computers create categories through patterns in data which are made useful by 'algorithmically transcoded ideas about the world' (2017: 37) called measurable types. Online we are not who we think we are, so much so that we are continuously assigned membership in temporary emergent categories based on our data. Using the example of web browsing patterns and other data generated as we participate in digital environments, Cheney-Lippold (2017: 5) demonstrates how our data is assigned categorical meaning without our direct participation, consent or knowledge. Measurable types of an 'unreliable or a high-cost worker' or a 'person at risk' is computationally calculated and we are then allocated into these categories that identify us.

Websites take the fact that a person spent a lot of time on a certain page as evidence of the fact that you are a middle-aged man and as proof that a campaign aimed at middle aged men centred content is succeeding. These algorithmic identifications of age and gender by search engines such as Google are not based on one's voluntary identification or physical identity documents but are inferences that have very little to do with the complexities of your lived experience. These algorithmic assignments and categorisations that displace your actual identity care very little whether you really are a certain gender so long as you ‘surf, purchase and act like that gender’ (2017: 7). These systems have the ability to distil digital features of what a measurable type for a ‘middle aged man’ or a ‘college educated person’ would look like in order to create an algorithmic caricature or a data based category that is used to identify the next person who comes along.

The perils of moving to a system where someone is identified solely based on how well they fit a measurable type (or interpreted through their datafied identity) are demonstrated by Cheney-Lippold’s worst case scenario of when a ‘terrorist’ stands in for a terrorist. Predator drones fatally misclassify a wedding party when they rely on a data-based assessment of a particular group of people based on their cell phone heat signatures and the fact that these signals were being emitted from remote geographic locations. This is a famous example of how a data-based representation or an information signature was used to displace contextual understanding to rely on metadata to compare against a pre-existing pattern of a ‘terrorist’ template. As revealed in the Snowden papers, in 2008, the modus operandi of US drone strikes went from targeted; based on the terror suspect’s name, voice or on the ground intelligence, to one based on terrorist ‘data signatures’. Data-driven definitions of terrorists were operationalised based on patterns or signatures in large scale data from cell phones and satellite imaging. Identities were assigned through such an algorithmically processed categorisation of metadata without consideration for contextual specificity, motivation or biographical profile. These signature strikes aimed to kill what looked like a ‘terrorist’ ended up unintentionally targeting wedding parties where individuals congregate outside of city centres with their cell phones creating heat signatures through data ‘as if’ it was a terrorist meeting[7]. In analysis and intervention at such a distant scale, insights are based on data signatures rather than the close analysis of units representing a fundamental shift.

Using measurable types, templates or data signatures to sift through vast amounts of data in real time to identify what we are looking for signifies a serious reconfiguration of administrative practice that has wide ranging implications. If data-based reconstitutions of an

---

[7] “In a more general sense, however, the obliterated wedding party may be the true signature strike of the post 9/11 era of American war-making, the strike that should, but never will, remind Americans that the war on terror was and remains, for others in distant lands, a war of terror, a fearsome creation to which we are conveniently blind” - Engelhardt (2013)

individual entity or a ‘datafied self’ stands in for the person, the validity of the judgements and categorisations made on basis of this algorithmic identity is entirely dependent on the quality of the data inputs and the digital proxies used. Social service interventions that determine how families are preventatively targeted for early offers of help designed based on algorithmic fit ‘on data’s terms’ are profoundly different to those that rely entirely on human input. As are ‘at risk’ categories generated through patterns in data and based on what ‘looks like’ risky, problematic behaviour without considering the voice of the welfare claimant and the singularities of their unique circumstances, motivations, or intentions. As these data driven technologies are rolled out on a larger scale, this means making decisions about support, substantiation[8] and risk prediction in child protection systems based on proxies, inferences and information signatures.

The systems and technologies that bureaucratic organisations use to categorise in the age of datafication take on a different order of abstraction; and focussing on the four themes below will help us investigate it further and develop a fine grained, empirical account.

2.1 Category Definitions and Proxies

Classificatory systems operate differently under datafication – there is a lot of reliance on automated category partitioning, on proxies and what is digitally observable. If the classificatory process in general creates abstracted, constructed groupings of people (or objects); in order to be legible and identifiable to newly datafied administrative procedures one has to be reducible to measurable types. Bracketing off what isn’t computationally relevant about a phenomenon in this manner means reducing it to discrete, tractable machine readable elements[9].

This has particular implications in areas of service delivery because someone deemed ‘at risk’ of lifelong welfare dependency or ‘at risk’ of child harm is compiled using a range of digital

---

[8] Substantiation is the statement from child protective services about the validity of a report of child maltreatment or abuse

[9] This process of prototyping flattens the continuum of people’s personal attributes and identities into ‘the discrete that marks the condition of possibility for computationality’ (Cheney Lippold 2017:53). It is only then that people and their problems can be ‘transformed’ into entities that such organisations can recognise and process.

proxies and observable information signals. If the model is looking to identify those who are, say, ‘vulnerable’ in order to target services, this would involve a kind of prototyping; with the data template for what looks like ‘vulnerability’ being assembled through a range of indicators and digital features. Various existing models have aggregated datafied elements like digital records of benefits claims, reported earnings, credit checks and household composition information to produce a risk assessment of ‘welfare dependency’ (König 2017). Unusual digital proxies like a change in the usage of local libraries in the service records of a child followed by a request to access counselling services have come to be read as a matter of concern and as digitally tallying against being ‘at risk of child harm’ (Malomo and Sena 2017)[10].

We have already seen how ‘at risk’ categories are constructions that reflect the anxieties of the welfare state and institutional norms and values encoded at the point of design. When patterns in data are parsed through for what *looks like* an ‘at risk’ situation are therefore ‘constructions about constructions’ (Cheney-Lippold 2017:45). The parameters for these categories are computationally generated and an empirical exploration of these classificatory processes would draw out the logics of equivalence and proxies for vulnerability/fraud that crucially has to be available in a digitally recombinant format.

Moreover, to understand the changes unleashed when welfare provision is based on assessments of fit into a ‘measurable type’ category, it is necessary to ask how such a new abstract category is translated into existing institutional routines – what explanations and meanings support category boundaries as reasonable to base administrative practice on? That is, how do those implementing these social service projects understand and legitimise the demarcation of ‘at risk’ from ‘normal’ and therefore not requiring intervention. This requires paying attention to the actual organisational practices surrounding these data driven

---

[10] This refers to the Integrated Data Model piloted by the Kent County Council to enable early intervention in child services (Malomo and Sena 2017). Documentation on the lessons learnt show that data points are inserted into a series of ‘at risk’ groups based on metadata scraped from the large textual corpus of social worker notes, and a larger dataset that includes information on library membership/usage, youth offending, educational outcomes and access to specialist child services. These datasets are matched with categories from the marketing social segmentation tool Mosaic which is designed to make consumer demographics, lifestyles, and preferences comprehensible to advertisers. Drawing on social context information on specific communities from this ‘off the shelf’, private classificatory tool and a pattern of irregular school attendance, it is possible to compile an information signature of children at risk of needing further attention from the council at an individualised level. Other indicators of concern in this case included the erratic use of local libraries followed by a request to access counselling services.

technologies and the situated manner in which people choose to use the output from these datafied models (Christin 2017).

2.2 Allocative harm and aggregative power

Within different forms of machine learning, gradient based optimisation models and other new generation big data methods, algorithms sift through data sets to discover 'emergent relationships attributes' (Nissenbaum 2009: 44). As a data reduction technique and method of analysis, algorithms aggregate data points into groups in order to somehow reflect the 'underlying structure' of the entities that the data represents. Critical scholars refer to the manner in which automated pattern detection techniques partition data points and organise and select clusters as a kind of 'promiscuous association[11]' (Sandvig 2018). This constitutes a form of profiling which seeks to assemble individuals into meaningful groups without their consent, for which representational exactness of the individual is irrelevant[12] (Hildebrandt 2011, Floridi 2012, Leese 2014). Therefore, there is a kind of 'aggregative power' (Ananny 2016) embedded in the manner in which people are 'semi-autonomously' sorted into groups by inferring associations.

Machine learning techniques can best be understood as teaching[13] the algorithm to replicate patterns and outcomes in the training dataset – it is a way of working backwards from the result to learn what the important factors are (that are correlated with the outcome) directly from the data. For instance, real world training datasets from the criminal justice system used to predict

---

[11] Individuals are placed into groups based on common patterns found in the data that might not be evident prior to the analysis; in fact these associations and patterns are unearthed in order to cluster and make sense of the data as part of the analysis. Instead of using statistical techniques to verify clearly formulated hypotheses about causal relationships between say demographic features and educational outcomes, this is a mode of inference where data mining applications run through the database looking for meaningful correlations between variables.

[12] In fact, Floridi argues that the individual's 'informational identity' is breached by meaning generated by algorithms that link the subject to others within a dataset (2011). Profiling by algorithms in this case is broadly defined 'as the construction or inference of patterns by means of data mining and . . . the application of the ensuing profiles to people whose data match with them' (Hildebrandt and Koops, 2010: 431).

[13] If a recipe is thought of as a list of steps leading to a defined goal, ML algorithms can be thought to start with the result of the recipe or the dish to work backwards to actually **learn** the whole recipe automatically.

likelihood of reoffending will have an overrepresentation of black and disadvantaged communities. An unsupervised (or supervised) learning model trying to simulate the conditions for reoffending will replicate this bias. In the case of, say, a model assessing financial risk of not repaying mortgages, people under 30 may systematically be put into a high risk credit pool because false positives based on patterns in the training dataset means that people are categorised into the wrong class[14]. Since the training dataset most probably did not have anyone under 30 successfully meeting mortgage obligations, it is likely that the model thinks that being under 30 is a good predictor of defaulting on loans.

This group-to-individual problem is a classificatory one, as group-based information, judgements and predictions about behaviour at the population level is translated into absolute predictions at the individual level (Mittelstadt et al. 2016, Binns 2019). When this happens, allocative harms occur as opportunities or resources are withheld from certain groups due to these models producing false positives based on judgements of group characteristics (Barocas et al. 2017). One can be labelled as 'at risk' of welfare fraud not based on individual behaviour or directly observable characteristics but information from other people in the large dataset, group traits and connections to others who display problematic behaviour (Newell and Marabelli, 2015:5). After pointing out how within current Data Governance law, these population-level relational effects are under-theorised and seen as incidental, Viljoen (2021) calls these 'horizontal data relations' that datafication puts people into; that have the potential to materialise unjust group-based relations like racism, sexism, and classism.

2.3 Data inferred classification and Context

Classification based on data technologies and big data analyses are crafted at a distance from human experience. Contextual details of circumstances are disregarded and insights are based on reductions and essentialisms and not on the close study of each unique case. Seeing society as categories of actuarial risk within the domain of child protection means looking for the highest priority risks that can be managed rather than trying to prevent all possible harms (Beaussier et al. 2016, Yeung 2018). This involves relying on the correlates of such high probability/high consequence risk and erasing everything that cannot be quantified by the risk-measurement model.

---

[14] Majority of ML and fairness researchers work on this problem of false allocation (see Sweeney 2013, Bolukbasi et al, Suresh and Guttag 2019)

Such contextual circumstances that get erased are relational aspects of the client-social worker interaction which are critical to the success of any welfare intervention. A problem is likely to be very narrowly defined so as to be computationally recognisable and by targeting behavioural traits that are correlated to risk and not the causal factors, such an approach can erase the contextual variations and the day to day aspects of the individual case. Categories have a flattening effect such that there is very little consideration for gradations of difference in the circumstances of parents who are for instance classified as having engaged in substance abuse (caregivers who admit to occasionally smoking marijuana after children have gone to sleep are put in the same category as caregivers who inject heroin several times a day and spend much of their time finding the means to do so – Gillingham 2015). It is alleged that such data based tools in welfare provision are oriented towards risk management and tend to 'amalgamate clients' circumstances into a series of factors to produce a risk score rather than see clients' contexts as a series of relationships, experiences and narratives which can be drawn upon to collaborate in problem solving and change agent activities' (Oak 2015).

Therefore, a technological system that tries to fit a welfare recipient into one category and not the other has always done so through a process of naturalisation, essentialisation and reification. For example, attaching labels like 'at risk of welfare fraud' is not as straightforward as affixing categories to objects that have a natural fixity (like labelling a chair or a table). We have seen how classification as a process reduces fluidity and gradations of difference in a person's situation into something that can fit into a fixed category.

Fraught as this process is with slippages and other boundary demarcation issues, this used to be based on a one-on-one interaction or at least the substantive discretion based judgement of a frontline caseworker. However, problems are amplified when category allocation comes to be based on automatically ***inferred*** behavioural characteristics compiled from digital proxies, information signatures and continuous online tracking of transactional trace data that does not necessarily relate to any specific individual (but refers to an 'abstraction of an individual' - Yeung (2018). Datafied 'at risk' marks a departure from earlier often administrative understandings of risk because it is an aggregation of data fragments and partial selections from multiple databases[15] or a category based on other categories.

2.4 Feedback

Furthermore, complex consequences follow as studies of high stakes algorithmic decision making show how the point of decision shifts within an automated system (Alkhatib and Bernstein 2019). Bureaucrats can act reflexively before making the decision – they can think about various incoming pieces of information, weigh up different evaluative criteria in their minds and then formulate a view. Whereas algorithms, based on how the implementation went, on receiving feedback, can at best retrain and act reflexively after the decision. Given the gravity of classificatory tasks that a fully datafied institutional practice would have to perform, it is necessary to ask if it is possible to feedback to the system and push back against its automated directives.

There is a distinct mismatch between mathematical optimisation and the categories that are generated from such an automated process and human reasoning and styles of semantic interpretation (Burrell 2016). In considering what datafication means in areas of public service delivery it is necessary therefore to focus on the ways in which data scientists and civil servants make use of these new categories and derive meanings from them. How do they navigate these new classificatory tasks and interact with these categories? If there is a disproportionate reliance on inferred characteristics, flawed training datasets and digital proxies; is there an awareness of the vulnerabilities of the system? Is there space to express disagreement with the decision support tool, override its recommendations and make changes to the risk outputs when situational details and information that cannot be captured by snapshot digital metrics come to light (Veale and Edwards 2018, Veale 2019)?

Processes of classification are intensified within datafication and take on unique features that are as yet under-researched in these administrative and social policy contexts. Measurable type models are increasingly reorganising service delivery and the logic of care. Following from

---

[15] Amoore uses the term ‘data derivatives’ to describe this way of repurposing and recombining digital traces looking for targeted behavioural characteristics where data collected from one context is made co-present with data from another.

this conceptual framework and the above theoretical discussion of datafication, the research questions for this doctoral project are specified in the next section.

3. Research Design

Welfare bureaucracies are undergoing a transformation with various administrative procedures that were once fully based on human inputs transitioning into data driven, computational systems. Focussing on classificatory processes that are now being made based on new kinds of linked, aggregated and non-traditional data draws attention to how decision making within welfare provision is undergoing profound change.

Allocative choices about who to target for preventative intervention, prioritise for services and identify as being at risk of welfare dependency[16] are beginning to be made on the basis of complex computational processes that draw on inferred behavioural characteristics and their digital proxies. There are many levels of interplay between these automated models and existing administrative (paper based) risk assessment tools, case work and bureaucratic routines. While private consultancies and the designers of these systems are proposing to deliberately 'redesign' entire 'service delivery models' (Ernst & Young 2019), it is still be to seen how integrated these new computational processes are within existing organisational routines. In order to understand how these processes of datafication are advancing in a real world context, empirical work is required. The above theoretical discussion has underscored the importance of the evaluative principles that are used to demarcate category boundaries and the meanings and explanations emerging around the use of new abstract categories in an existing institutional context. In order to empirically study how these computational processes are being used to sort and profile individuals, households and communities, I will focus on answering the following research questions.

Research Questions

1. How are social categories being developed and translated into data categories?
2. How are administrative concepts like 'at risk' operationalised in algorithmic infrastructures for making decisions about welfare provision?
3. What discourses within bureaucracies are used to justify the implementation of processes referred to in 1 and 2?
4. As a result of these changes, has the role of the human frontline worker shrunk? Is there now a much diminished scope for their expertise? How do they see their role in conjunction with these processes of datafication?

---

[16] Or at risk of 'welfare fraud' as a recipient of overpayments and needing to be censured.

RQ 1 aims to explore how social categories are captured as inputs into predictive analytics or risk scoring models[17] by gathering accounts of the new parameters of the category and the digital measures for them. For example, social services and welfare professionals often use social categories like ‘family in acute stress’, ‘absent parenting’, ‘anti-social behaviour’ and ‘family’s social integration’. Within older administrative assessments in child protection, there was a requirement to note the ‘social presentation’ or ‘self-care skills’ of the child. Is it possible to find digital proxies for these and if so, how were they arrived upon? ‘Anti-social behaviour’ (coded, reified and naturalised as variable ‘ASB’) is scored in many child protection models and is often presented in its business case documentation as one of its main advantages. Child care payment fraud detection models quantify family and undeclared extended family relationships based on administrative data alone to verify if the child care subsidy for the same person is being fraudulently pocketed by two households with different last names. Peer networks are quantified and captured as data categories; as are various kinds of vulnerability.

Bureaucracies require complex operational concepts like ‘at risk’ to target and customise service delivery. Many datafied models aim to produce a complex measure of risk or reduce it to a score that can be used to comparatively assess who to prioritise. Old administrative categories are being re-operationalised and taking a different form as well. Insights on the changes wrought by the datafied model can be elicited by leading a structured discussion on the component factors of risk. RQ 2 aims to gather empirical detail on how these are changing in the datafied era.

While RQ 1 and 2 refer to the technical procedures and descriptions of what seemed to be happening to those close to the process, RQ 3 seeks to understand the justificatory discourses, rationales and explanations that accompany these data processes. As emergent technological processes, the discourses surrounding and promoting them have consequences for how the relationship between the data and the problematic behavior they are supposed to signify is framed and narrated (Van Dijck and Poell 2016). Are the datafied processes framed as having a one to one equivalence between digital behaviours and real (offline) problematic behaviours?

---

[17] The aim is not to unearth every perception about social roles, positions or stereotypes that feeds into the assessment process but to start with the datafied model and lead a guided discussion on how relevant inputs were chosen, suitability of proxies and whether previously important social categories are still captured in the model.

RQ 4 refers to the administrative procedures and tasks now being made redundant and the self-reported reflections on the trade-offs from those implementing and interacting with these systems. There could be various forms of contextual and experiential knowledge, situational judgement or practice wisdom that the case worker or social worker holds that is now hard to make count in the datafied process. Reflections on how they see their role in conjunction with these processes of datafication are as important to map as the practices (RQ 1 and RQ 2) and discourses (RQ 3) of datafication.

By intensively researching a series of case studies and using a qualitative, multi method approach, I will provide convincing answers to the above descriptive questions. I use expert interviews and document analysis to develop two in-depth case studies:

1. Datafied 'at risk' modelling in Child protection systems within Local Government
2. Various uses of data and advanced analytics in the 'digital transformation' unit of a large Central government department.

Previous studies that have tackled classificatory aspects of data driven technologies have focused their empirical investigations on how sorting and clustering can produce problems of visibility and bias within social media (Bechmann and Bowker 2019) and in the context of financial credit scoring systems (Fourcade and Healy 2013, Citron and Pasquale 2014). There are very few research projects that have mapped in a granular manner the roll out of these data technologies in the public sector and within welfare provision. There are a few studies (Redden 2015, Dencik et al. 2018, Andrejevic et al. 2020, Redden et al. 2022, Currie et al. 2023) that have begun to highlight the public private partnerships and the subsequent valourising of economic logic that this phase of datafication enables. O'Neil (2017) describes mass dismissals carried out after public school teaching performance was based on data driven models. She draws attention to how those implementing these customised systems built by profit driven, proprietary consulting firms did not understand how it worked, yet still enforced its outputs. Eubanks (2018) documents instances of 'technological redlining' and profiling in various levels of government in the US within the fields of welfare eligibility, housing programs and child and family services (see Petty et al. (2018) for a similar qualitative study but with a different geographic focus). This research focusses on the lived experiences of welfare recipients who are most affected, with the perspectives of the 'data workers' closest to the changes unleashed by processes of datafication - civil servants and data science professionals within government - being elided over. Therefore, these research questions haven't been investigated before and addressing them by drawing on the experiences of those who design and implement these systems will provide fresh insight into how new kinds of classificatory practices[18] are being spawned by new data technologies within professional welfare work.

This research contributes to the growing field of Critical Data Studies which challenges the objectivity, rational authority, and ontological superiority of data by making visible the socio-technical conditions of its production and use. Scholars have underscored the futility of trying to understand data driven technologies as a purely technical pursuit with empirical research focussing on explicating just its technical aspects and deterministically taking its efficacy for granted (Christin 2017, 2020). By developing organisational case studies that elicit contextual explanations of how datafication is unfolding in practice, I am able to contribute an empirical account of how institutional categories are automated and datafied while tracing the manner in which these transformations are contingent on institutional, political and social factors. By highlighting the political and ideological threads underpinning data infrastructures, I redress the dearth of research that generates a contextually embodied and situated understanding of data. The local organisational realities that shape the roll out, uptake and production of data driven technologies have received less attention from Critical Data scholars and my research fits into this lacuna.

After outlining a conceptual framework focussed on classificatory processes and the issues raised when they are exacerbated by datafication, I pose four research questions. Decision making based on datafied models involve problems of category definition based on digital proxies (section 2.1) and allocative harm (section 2.2) when groups of people are systematically denied opportunities or resources due to automated classificatory mistakes. Classification based on metadata at a distant point of view violates contextual integrity and raises questions about automated behavioural characteristics that can be compiled to classify those who digitally look like they are ‘at risk’ (section 2.3). How data scientists and civil servants make use of new categories, derive meanings from them and whether they have the space to override its recommendations has consequences for feeding back to the system (section 2.4).

---

[18] Furthermore, classic studies of classification like Bowker and Star (2000) adopt the methodological strategy of identifying new categories and interviewing those affected by it. They ask the nursing staff whose work tasks are being classified and formalised into the Nursing Interventions Classification (NIC) system what they thought about them. Mary Douglas’s institutional theory around classification (1986) was based on interviews and ethnographic methods to understand how people wove meaning and symbols around categories. I aim to build on these approaches.

Drawing on the theoretical and conceptual framework outlined in this chapter, this thesis develops in-depth case studies of datafied child protection and welfare provision through fieldwork conducted between 2020 to 2021. The empirical chapters that follow build on one another, and together advance a detailed look at the expansion of datafication in sensitive policy fields in the UK and subsequent tensions. The presentation of empirical material and key findings in each chapter is interwoven with a discussion of how this research contributes to the theoretical debates in the literature with only the salient contributions reiterated in the conclusion chapter. Chapters 3 to 4 presents findings from the local government case study and Chapter 5 and 6 draws on the central government case study. Chapter 7 focusses on the justificatory discourses (RQ 3) and draws on empirical material from both case studies as does the final empirical chapter on Reflexive Dataism (Chapter 8). The next chapter sets out the methodological approach and some of the practical challenges I faced while operationalising these research questions and conducting qualitative fieldwork amongst insular, expert communities during a pandemic.

Chapter 2

Methodological Approach

In the previous chapter, I set out the core interest of this study – the little understood process through which data driven tools are being used to sort, rank and classify welfare recipients. The research questions seek to empirically map the justificatory discourses and practices of datafication by investigating how social categories are being developed and translated into data categories. I aim to understand how administrative concepts like ‘at risk’ are being operationalised in algorithmic infrastructures and document opinions and reflections on the human frontline worker’s role and the scope for their expertise in the now datafied administrative system. This chapter[19] describes my case selection criteria, how I developed the methodological approach to research ‘black boxed’, inscrutable algorithmic systems that are the focus of this study and some reflections on carrying out fieldwork under the severe constraints of the pandemic.

In order to research the expansion of processes of datafication into sensitive policy fields like child protection, welfare and social care administration, I built two case studies at different levels of government using a qualitative approach. Assurances of anonymity were provided before interviewees agreed to participate – so the local government case of ‘at risk’ modelling that combines statutory and enforcement data held by the council and the metropolitan police constabulary is referred to as Project Vision C. The central government department is referred to as Department A while reporting from interviews. I consciously adopted a comparative case study approach as it becomes possible to see how processes of datafication are unfolding in institutional contexts that are similar but reveal different aspects of the growth of data driven technologies and the organisational practices that surround them.

Case Study 1: Project Vision C
Case Study 2: Department A

Cases were screened for inclusion in the research based on the length of engagement with a data driven technology. The data pilot should have been running for long enough for there to

---

[19] As ‘data’ is the object of this study, in this chapter, please note that as much as possible I have tried to avoid confusion between data collection for the thesis and ‘data driven technologies’ that combine various forms of data records by referring to fieldwork in terms of ‘collecting empirical material’.

be observable changes – at least 2 years. These cases are exemplifying cases[20] (Yin 2014), typical of the kind of data innovation proposed in the social policy sector. The majority of data driven pilots that have been studied in the literature involve models developed for government by the private sector. This has led to theories about proprietary systems that are not understood by those in the organisation yet whose automated outputs are nonetheless enforced by deferring to the logic of dataism. It was important that each of the cases had a prominent in-house analytics and data capacity, so that I could explore whether departments with the capability to manage their data technologies and digital transformation mandates talk to frontline workers to get their input and practice wisdom before automating processes.

2.1 Entering the field

Answering the first two research questions about how social categories are being developed and translated into data categories and how administrative concepts like 'at risk' are operationalised requires eliciting technical details about the modelling process within respective case studies. This raised the dilemma of whether interview questions can be used as a window to an empirical world where the technical complexity of these data driven technologies make it hard for respondents to understand and articulate the mechanisms of their operation to me as a researcher who is a disciplinary outsider.

In response to concerns about the inscrutability of data driven systems that are often cast as 'black boxed', opaque and impervious to empirical enquiry, there is a growing number of methodological approaches designed to produce insights into their nature and work by teasing apart the assumptions, intent and objectives programmed into these systems. These methods include reverse engineering code[21] (Diakopoulos 2014), performing algorithm audits[22]

---

[20] They are not extreme cases or unusual cases used to challenge theoretical propositions. These data driven pilots fall within key social policy portfolios and employ similar methods like data matching, data linking, predictive analytics and large scale aggregation of dynamic, real time data streams. These are the kinds of 'data for good' social policy projects that are gaining salience; and are widely referred to as worth replicating in other domains.

[21] Where carefully selected dummy data is input into models to see what the relevant output is under various scenarios.

[22] This growing discipline of algorithmic audits investigates how the system operates in the real world and whether it is behaving a different way than was expected with regard to procedural regularity – i.e. whether the same procedure is applied to everyone including those in protected classes such as racial and ethnic minorities. It uses the following designs: 1) code audit, (2) non-invasive user audit, (3)

(Sandvig; et al. 2014, Bennett Moses and Koker 2017) or auto-ethnographies of translating tasks into code (Ziewitz 2011). Among these, a promising way to elicit details about design choices is by interviewing designers of the project and those at critical points in the data lifecycle about the inputs and outputs of the model, and to lead a discussion on how it weights and preferences some criteria (Kitchin 2017).

Focusing purely on the process rather than the goal or consequences of the model[23], regardless of how complex the system is, there are inherently two openings that enable lines of enquiry: inputs and outputs. By examining what datasets are fed into the respective analytics model and what output is produced, it is possible to ask respondents questions that open up a discussion about the criteria by which the model selects inputs for processing, what the activation functions (in the case of a machine learning model) are and what outputs are produced.

Kitchin (2017) and Diakopoulos and Koliska (2017) suggest that during the design phase which begins by developing goals and requirements of the data analytics project or pilot, once the task is defined and outcome variable chosen, most projects produce ‘pseudo code’. Pseudo code is a series of steps[24] or a process flow chart detailing how the task is translated into the model and the input into the outcome. By following these lines of enquiry and asking about artefacts and rulesets like the ‘pseudo code’, it is possible to understand technical details of how the proxies for social categories are chosen and bureaucratic concepts operationalised in complex models. Retrospective questions about how they framed objectives and made choices can be addressed to data scientists who built the system (internal or consultant) and the senior welfare professionals who were involved in design meetings and part of the decision to introduce these changes into the department.

In addition to these methodological tools from the ‘auditing algorithms’ literature that unpack technical details, it is necessary to note that these data driven projects are being introduced into

---

scraping audit, (4) sock puppet audit, and (5) collaborative or crowdsourced audit. This approach has been particularly successful in identifying price discrimination and bias on e-commerce websites.

[23] Most algorithmic audit tools are focused on consequences or effects of the algorithm such as bias or privacy.

[24] Pseudo code then has to be converted into actual code. This step falls within the planning stage of the project.

institutional contexts. Organisational Sociology[25] offers approaches to build a picture of the workflow even when you don't have complete access to the entire 'shop floor'. This is by interviewing key people within different work groups and asking them about others in the chain (Bryman 2013) and the decision tasks that are attached to each stage of the workflow. In this way, one can build an idea of the organisational chart, the workflow and triangulate from one respondent what the other does and thus build up a fuller picture through a snowballing process. Combining these approaches from the 'auditing algorithms' literature and Organisational Sociology, I identified the key decision tasks at various points in the workflow that was impacted by the introduction of the datafied model. Therefore, the overall case study exploration strategy was to trace the workflow and develop questions that ask data scientists about design choices and practitioners or frontline users of the datafied system about how they interact with output from the model.

I passed my upgrade at the end of 2019 and received ethical clearance for fieldwork at the end of February 2020 when the pandemic emerged as a global crisis with the national lockdown forcing the government teams I needed to access into remote work. This meant a significant disruption to the methods and object of study; that had consequences for how I would develop rich contextual descriptions of the layered meanings attached to the technologies and differentiate between the nuanced negotiations made by data workers when wrestling with competing interests. Traditional workplace ethnographies and face to face interviews were no longer possible. Work from home guidelines and legal restrictions on travel meant that my respondents as an elite population became more inaccessible than ever[26].

I had to develop alternative ways of engaging deeply with the field and started systematically collecting documentary evidence to build my case studies. Similar to pseudo-code, I started collecting process flowcharts of the modelling that the key data scientists in my cases where presenting at public forums like the fortnightly 'Data Bites' sessions that were organised by

---

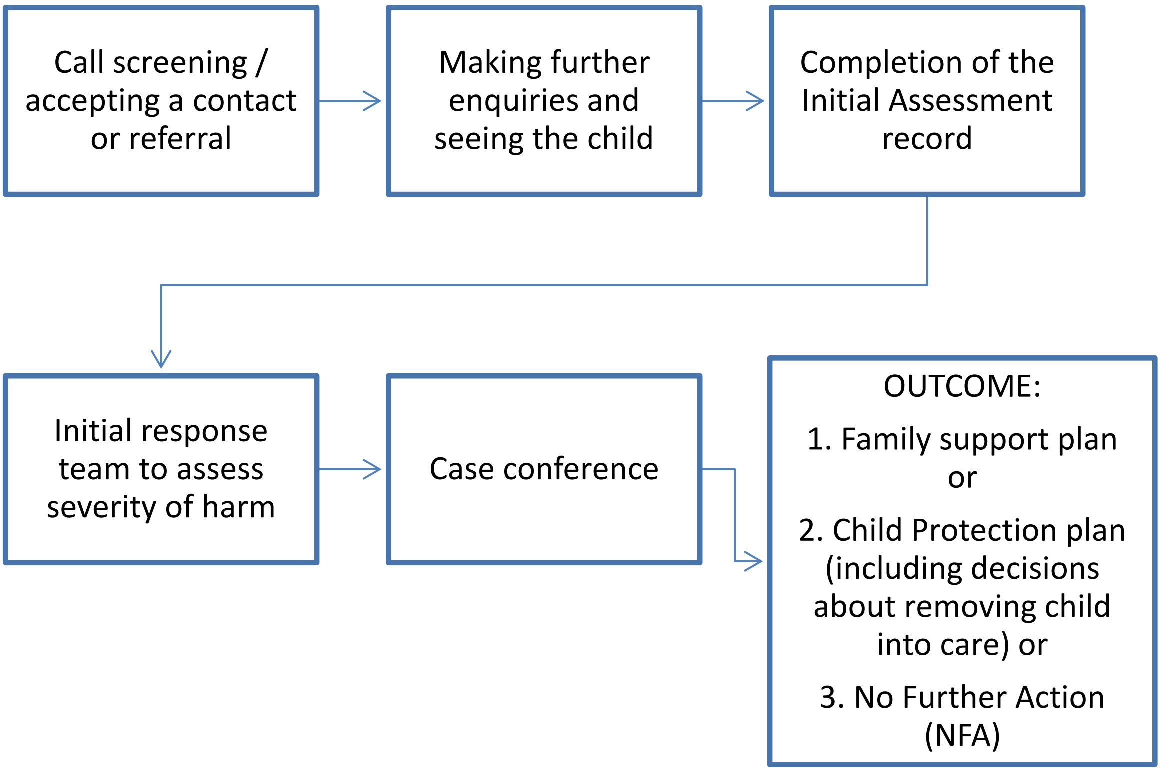
[25] Early studies of the introduction of IT systems and the subsequent changes in work roles, automation and the overall organisation of work (Kraemer and King 1986, Orlikowski 1989, Orlikowski and Baroudi 1991, Malone et al. 1994, Zuboff 1988) drew on the distinctive work place case study research methods from employment relations research (McGovern 2012, McGovern and Alburez-Gutierrez 2017). Reviewing the methodological practices in these studies, labour process sociologists have developed systematic approaches that trace the workflow even with limited access.

[26] This is despite the fact that I had at that point spent months trying to build relevant contacts for fieldwork access

the Institute for Government and archived on YouTube and presentations to local government Overview and Scrutiny Committees. Drawing on my familiarity with coding, I began reviewing the public GitHub repository of the central government department. As part of the service standard instituted by Government Digital Services (GDS), they are legally required to deposit a version of their key prototypes. Based on such a technical outline of the workflow, I reformulated my interview strategy to ask open ended questions about the inputs and outputs of the model in order to elicit an account of how the model parameters, predictive variables and proxies were chosen.

Flowcharts such as the one below of the pre-datafication child protection workflow were displayed on the screen during the course of interview, so that the respondent can visually identify which decision tasks are changed when they rely on the datafied model. In some instances, alternate flowcharts were produced by the interviewee to correct my understanding.

Figure 1: Pre-datafication child protection workflow (simplified)



Government teams took a long time to adapt to remote work. The hardest hit were respondents in local authorities who would have been part of my first case study who were at that point working overtime beyond normal business hours and swamped with responding to a rapidly unfolding situation. I was told that if things get worse and the infection spreads in care homes that their social care teams have a statutory responsibility towards, plans were for senior staff in the local authority to provide care i.e. Directors of Social Services providing home care. It was challenging to attempt to talk to respondents over zoom, to build rapport, explain my research aims and recruit them for a full recorded interview amid the stress of the pandemic[27]. By around September 2020, some of the major tech fests and public sector digital transformation conferences restarted and were being conducted online on Zoom. The central government department started participating in some of these regional tech conferences such as the Leeds digital festival by reporting on how they have adapted to the pandemic. I resolved the initial access issues by asking engaged questions at the end of presentations in such a community of their peers which meant that they had to respond. Drawing on my participation in previous in-person hackathons and public facing organizational learning workshops that this particular department had conducted since 2018, I was able to frame my research aims in terms of 'change management' using issues that I knew were of concern to the technologists in the department. As data specialists working in what they perceived to be a department with a 'legacy' culture, sympathising with their challenges and drawing on my background working in government, opened up conversations. Before these events restarted, in the initial emails sent introducing myself and soliciting participation as well in the verbal pitch, descriptions of my research aims were kept necessarily brief:

> *I aim to uncover contextual and organisational factors that lead to digital innovations being embedded and scaled up. I am interested in the manner in which data analytics and decision making based on advanced modelling interacts with the organisational culture and existing working practices of social care teams within government.*

As technologists facing some opposition in a bureaucratic context, they were interested in these organisational culture issues and the answers that my research project would find. They were keen to talk about the cultural changes needed in a government department to sustain digital projects. I therefore started interviews with these technical issues, then used this interest to get

---

[27] The third lockdown in January 2021 made it difficult to sustain conversations with initial gatekeepers and a situation arose where no one would respond to emails or follow up requests. The initial interviews that were carried out varied in quality and it was not until Summer 2021 that I was able to secure access agreements from two organisations that agreed to let me conduct repeated interviews with the designers and users of the data technology in order to form a fuller picture and build an in-depth case study.

them to open up about the analytics methods they use at the end. In one instance, a key data group manager provided extensive access after I responded to the team management challenges posed in his talk about remote work where he found that millennials had a comfort level with digital communications while senior managers found the pandemic unworkable causing problems and conflict. I helped design a small research project that helped him explore this further and in exchange he secured introductions and provided access. The lead data scientists at these sorts of events would typically talk about accelerating the digitization of key benefit lines which extended from changing the personal independence payments (PIP) end to end, digitising health assessment services and access to work services. These descriptions and preliminary analysis of their data ecosystem from material presented in tech show-and-tell events over the last three years, meant that in the full recorded interview that followed, I was able to ask specific questions that assumed an extensive knowledge of their systems.

Even though it took a while to gather sufficient interviews from each case study, through this recruitment strategy I conducted interviewees in two phases with shorter interviews where they describe the technical project to me (17 informal/background interviews) and 20 in-depth fully recorded interviews. The shorter interviews where I heard them talk about the risk alert and automated outputs from the model was limiting in terms of how I could explore how they activate discourses around social categories or get them to reflect at length. However, I was able to gather information on technological rationales and form a background picture. 20 in-depth interviews were then conducted with 8 interviews corresponding to the local government case study (Project Vision C) and 14 interviews forming the basis of the second case study (Department A). Two respondents asked for the interview not to be recorded and two asked for the transcript that they could review before my use. Using a snowballing approach, I was able to identify and engage with current employees and former employees who are freer to speak. Anonymised details of those interviewed for each case study has been provided in Appendix 2.

Table: 1 Breakdown of background interviews and details of negotiating initial access

	Approached for initial fieldwork	**Interviewed or conducted informal discussions with**	**Notes**
Case Study 1	London Borough of Tower Hamlets	2 - Principal Social Worker & Head of Social Work Academy, Head of Early Help (Interim), Children and Culture	Declined to allow access for full fieldwork and further interviews
	Kent County Council	1	Declined to allow access for full fieldwork and further interviews
	Barking and Dagenham	5 - Head of Insights and Innovation, Behavioural Science lead, Data Ethics and Policy Officer, 2 external consultants	Did not respond after initial contact
Analytics consultancy firms	Ernst and Young staff member, Xantura CEO		
Case Study 2	Government Digital Service (GDS) data scientist		
	Department A staff at public facing events in 2018, 2020 and 2021	Overall interacted with over 20 members of staff including frontline staff – informal discussions and access to internal github repository	

Other documentary sources for Department A included an archive of articles by Chief Information Officers in Computer Weekly, Weeknotes written by data scientists reflecting on their work, and extensive blogs maintained by the department that detailed the evolution of the data capabilities and the digital website. A recent autobiographical book released by the former Minister for Welfare Reform had a compendium of key policy documents such as white papers, NAO reviews and internal documents relating to the welfare portfolio.

2.2 Constraints of Zoom interviews: Conducting fieldwork during a pandemic

> ***AT**: That's, that's it. Any other lessons learned or any other ways in which data has been useful to your work that we haven't covered already?*
> ***Interviewee 8**: Yeah, sorry. You got quite a lot of feedback there. Again, I think the question was, is there anything we haven't covered? which we haven't covered already? Was it? Nod for yes, shake your head for no! There's nothing more for me, I don't think.*
>
> ***Interviewee 15**: I'm just conscious that we've got a minute and 34 left on this zoom call. So should we leave and come back? And I know we had till about 11:45 am. But then that just gives us a bit of a buffer to wrap up any outstanding questions that Asha has. Is that alright?*
> ***AT**: Yeah,*
> ***Interviewee 15**: Lets leave this call and come back in. [pause as we log out and log back into zoom]*
> ***AT**: Okay, I just have about two more questions. And then we can wrap up*

Beyond access issues, the conduct of research through remote zoom interviews meant that I had to find different ways to ask questions that require more trust from the respondent. Scenarios were used to get them to think about aspects of their work that might spontaneously come up in a conversation in a physical space where I can see the model they are talking about or pointing to on their computer desktop. As the focus of the research narrowed, I had to find ways to gather contextual information that I can’t directly observe anymore. These changes in the data collection process involve more than just adapting the topic guide for online use and communication struggles like the ones displayed in the above quotations from two separate interviews.

In addition to the technical details on the mechanics of the data model, it is necessary to elicit an account of what the process seemed to be doing to those close to it. While this was not fully possible for Department A, with the local government case study, semi-structured interviews with the designers of the model (data scientist) were followed with interviews with users of the model and generalist program managers (council staff) with oversight of the process such as administrators of existing council teams who have seen their outcomes and referral criteria changing. By contrasting responses from them with those who are closer to the implementation of the model and have to make assessments based on the risk alerts enabled me to get a rounded view. Discourses that surround and promote the model and the justifications that explain the shift, the need for it, and why they think it is working need not be mirrored by the experiences of those on the ground. Triangulating responses from those at different levels of the organisation revealed multiple perspectives that can be supplemented by the documentary material.

In order to achieve depth of analysis and build a robust body of evidence from different sources, interviews were coupled with document analysis (Bennett and George 2005). Relevant documents include business case proposals, presentation packages framing the successes of the project and privacy impact assessments undertaken by councils that are now a legal requirement under GDPR and tender/procurement documents. The framing and rhetoric of these projects and the clear causal language found in various policy documents revolves around economic rationales and the merits of early intervention to save costlier services in the future. By adding documents to the corpus of data I collected, it thus became possible to trace coherent themes in the discourse when text segments, expressions and explanations are reused or only slightly reformulated.

A key outcome of the argumentative turn in Critical Policy Studies is this close attention to language and narrative in how the problem and the policy solution is framed (Garvin and Eyles 2001, Fischer 2003, Fischer 2015). However, in seeking to build a rounded picture of the case with both interviews and documents, it is necessary to acknowledge that official documents present a well-crafted, polished narrative produced to promote and garner support for the data pilot. Many relevant justifications and ‘argumentative threads’ (Majone 1989) are buried in off-the record briefings, meeting minutes, and political speeches that are inaccessible to this analysis. There is much existing methodological work on interpretative policy analysis that shows how it would be naïve to treat these documents as records of decisions made or as a direct window to retrospectively reconstruct the empirical conditions surrounding the introduction of the data pilot (Prior et al. 2012, Jones 2013). My work builds on these approaches and extends this method because the data project still has to be explained in appropriate terms for there to be support and this narrative posturing is important. By treating policy documents as an active attempt to shape understandings and optimistically generate expectations for the pilot’s success, it is possible to address both its communicative intention, contribute to analysis and even provide prompts to dig behind statements made during semi-structured interviews.

I used thematic analysis to inductively code and work with data from both sources. Requiring a high degree of involvement and interpretation, this is a widely used analytical tool to identify and describe both implicit and explicit ideas in the data as guided by research questions (Attride-Stirling 2001, Braun and Clarke 2006, Guest et al. 2011, Ezzy 2013). Thematic analysis 'starts to identify or examine the underlying ideas, assumptions, and conceptualisations - and ideologies - that are theorised as shaping or informing the semantic content of the data' (Braun and Clarke 2006: 89).

Interviews were transcribed verbatim and after familiarising myself with the textual corpus (of both documents and interview transcripts), I produced initial codes based on patterns in the text in a grounded, inductive way. After collating individual codes, I searched for themes. For example, the proxies of 'at risk', inferred problematic behavioural characteristics and datafied categories that emerged from the open ended questions about the modelling process were each themes with subthemes such as disagreements around how to define category boundaries (comprising of statements about how these disagreements were overcome).

A theme that linked research questions 3 and 4 was explanations that draw on the pre-datafied work tasks – interviewees justified these automated categories as being reasonable to base administrative practice on by drawing on and activating analogies[28] to how pre-datafied tasks were performed. In this manner codes and themes were developed organically from the data and while inductively working with the text.

2.3 Conclusion: Reflections on methodological choices and ethical trade-offs

Being forced into remote mode significantly influenced the methodological design of this study. Design choices were also guided by ethical considerations and I sought to minimize, mitigate and manage any potential risk to participants at every juncture. In planning the methods, it was imperative that no one who agreed to be interviewed would lose their jobs as

---

[28] For example a rationale that is often offered in pitches to scale up the child protection data project is that it is 'merely collating records' from various agencies and thereby saving the time poor social worker's energy. In the case of Department A, I heard that a chat bot that uses natural language processing to handle queries from citizens on the website was basically a way of producing a script for a call centre operator. Automated keyword detection techniques were justified by comparison to how a junior staff member uses the script to guide the claimant through a set of options. If the junior staff member can see that they asked a question and got this advice but then took that wrong action, 'how can we improve the advice so they take the correct action?' (Interview 2). Producing a chatbot is very different to writing a new help guide to correct a misunderstanding – however, the radical change activated when chatbots are used to deal with vulnerable welfare claimants was justified by saying that it mirrors offline workflows.

a result of participation in this study. To this end, a strict confidentiality protocol was worked up and assurances of anonymity were provided.

I promised each respondent that I would refer to them using pseudonyms such as ‘Manager 1’ or ‘Data Scientist 2’ and that insights from the interview would be incorporated into my analysis of the broader issues in the field of child safeguarding and welfare provision so that they are not directly identifiable. Informed consent was sought on a one-to-one basis and I reiterated that participation was voluntary, that they can opt out any time they chose or ask to turn the recorder off midway through an interview. Potential participants were briefed on the goals of the research individually and it was made clear that as an independent researcher, I do not represent the interests of the manager or the Tech Company running the data pilot that referred me to them. Moreover, the content of their interview would be treated with strict confidentiality and not be relayed back to their line managers. In the case of senior managers or someone higher up the organisational ladder, there is a risk that revealing their exact role would compromise their identity. At the same time, it would be meaningless to research the advance of datafication across a government department without talking to the relevant ‘Head of Role for Data Science’, ‘Head of Role for Agile’ and so on. In reporting on interview data from these individuals, a deliberate choice was made to use ambiguous role titles such as ‘a senior manager’ rather than list their actual job title. A few of these senior civil servants, who despite being named speakers at high profile digital transformation conferences in the public domain, were hesitant to proceed with the detailed parts of the interview. In these cases, I offered to provide the complete transcript to them for review by others in the department and at one point even offered to only use the amended revised transcript for my analysis[29].

At the start of each interview, I meticulously explained that it was their prerogative to ask me to not record the conversation, to not answer a question they were uncomfortable with or withdraw from the study at a later date despite allowing me to conduct and record the interview. Several respondents asked for the recording to be turned off at a certain point in the interview in which case, I would later reconstruct the broader themes of their answers from memory. Often, I would be able to come up with another question at a later date that elicited the same information using an example they were confident to talk about as I found mentions of an

---

[29] While I did provide the final typed up transcript, this offer to edit the unprompted responses to my semi structured questions in a document that was 16 pages long was (thankfully) not taken up.

equivalent data use case in either the documentary material or found a department podcast segment of them discussing the same point to their media team.

Therefore, in order to meet the ethical obligation that the risks of participating in the research is minimal especially for junior data scientists, considerable time was spent in preparing consent forms, and providing clear information about the proposed research while seeking informed consent. The reputational risks to the participating data scientists and civil servants are negligible when compared to the disproportionate impact of automated decision making on vulnerable welfare recipients. Since the benefits of the research far outweigh the potential harms, while planning for the interview-based case study design, ethical concerns I gave considerable thought to were around securing informed consent. Methodologically, I planned options if a gatekeeper were to object to me and block access by approaching senior respondents later on in the data collection process. I also reviewed the dataset on an ongoing basis over the course of the two year long fieldwork period to see if a fuller picture of the case study could be formed by identifying and interviewing former members of the data project who now work for another organization and are freer to speak.

Having clarified and elaborated on the range of methods I used to research data driven tools in an organisational context, I conclude by providing an account for design and fieldwork choices by addressing a few points on methodological reflexivity. In order to address issues of rigor and methodological robustness, I will briefly list the value of the case study approach I adopt in studying a complex phenomenon in context (while using elite semi structured interviews and drawing on a systematic collection of relevant policy documents and their analysis).

Aiming to generate an intensive study that highlights the detail, richness, completeness and within-case variance, a case study approach ‘closes in’ on real-life situations and facilitates the direct exploration of ‘phenomena as they unfold in practice’ (Flyvberg 2011). In order to understand the introduction of these new technologies into various government departments, it is necessary to elicit contextual explanations from the perspective of participants in these policy communities[30]. There is an emerging body of literature on algorithmic decision making that employs experimental study design to investigate bias and automated risk assessments (Green

---

[30] These include descriptions of the existing administrative procedures being replaced, imagined benefits, policy rationales and lessons learnt from the rollout of these technologies.

and Chen 2019). In contrast to an experimental approach that deliberately separates a phenomenon from its context and simulates environmental conditions to study causal relationships, the in-depth case study method is particularly useful when the boundaries between the phenomenon and its context are not clearly evident (Stake 2006, Simons 2009, Yin 2014). In both case studies data driven processes are at a pilot or mature implementation stage that has been introduced into an existing work flow and bureaucratic culture. Tracing the interaction between the data driven technology and the context it has been embedded into (as well as the other relevant factors at work) is crucial to the objectives and structure of the research problematic. In addition to a thorough contextual examination, Gummesson (1988) argues that an important advantage of case study research is the opportunity for a holistic view of the process:

> "The detailed observations entailed in the case study method enable us to study many different aspects, examine them in relation to each other, view the process within its total environment and also use the researchers' capacity for 'verstehen'".

Bureaucrats and data scientists employed in government digital teams hold specialist knowledge of administrative procedures and the details of policy implementation that would otherwise be 'off-stage', unrecorded and inaccessible to anyone not directly involved (Lilleker 2003, Smith 2006). They are the only source of direct information on their own activities, day-to-day practices and the scope and context of their roles. In order to uncover details about the roll out of these technologies, it is therefore necessary to make contact with these civil servants. In this regard, semi structured, elite interviews are an effective tool to gather rich and detailed information through a guided conversation in which the researcher carefully listens 'so as to hear the meaning' of what is being conveyed (Rubin and Rubin 2011:6, Kvale 1996). Using main, probing and follow up questions it is possible to get "a fine textured understanding of beliefs, attitudes, values and motivations in relation to the behaviours of people in particular social contexts" (Bauer and Gaskell 2000: 39). Despite only being able to interview a relatively small number of respondents, the aim of elite interviews is to draw on the reflections of key players within critical points in the organization in order to get their account of how these datafied processes are being implemented and why. By definition, elites are less accessible, but the value of the material is in gathering explanations for major decisions that helped shape the outcome of events and soliciting 'behind the scenes' interpretation of information that is not recorded elsewhere or not yet available for public release. Such respondents also can provide access to others involved in aspects of government that are hidden from public view through a snowball effect. Methodologically, advice on handling elite interviews with respondents who are often the only source of information and specialist knowledge on government issues that they are deeply involved in, relates to mitigating unequal power relationships and ensuring that interview data is systematically triangulated with all available documentary evidence and facts are corroborated thoroughly (Richards 1996, Lilleker 2003, Smith 2006).

While there are several advantages to using semi-structured interviews, the choice of method in this case is informed by the data collection needs of this project. The purpose of interview based research is not to count the frequency of opinions or people but to understand the range of opinions, and the variety of views and representations of the issue (Bauer and Gaskell 2000: 44). The interpretative and explorative nature of the research questions, therefore, necessitates data collection using open ended questions as opposed to the fixed choices that surveys allow. In conducting such interviews, I was able to adapt as the conversation progressed and pursue promising topics and lines of enquiry as they emerged. This made it possible to gather unexpected information that other more structured research instruments preclude (Berger 1998).

Since methodological choices determine the nature of data that can be collected and the kinds of questions that can be answered, it is important to reflect on alternate methodological approaches and the limitations of the adopted data gathering strategy. Within the broad suite of qualitative methods, focus groups are widely used to elicit shared assumptions and ideologies on a topic. However, I seek to elicit rationales for the datafied model and automated category definitions and retrospectively reconstruct the thinking and storylines activated during the design phase of the data driven tech project. This requires the ability to lead respondents individually through the decision making process from the past rather than generate a snapshot of current opinions as produced by a focus group interview.

An obvious alternate approach would be an ethnographic one where there would be the opportunity to participate in meetings, watch what is happening on the office floor on a regular basis and then cross check main themes from the participant observation against policy documents, minutes and reports[31]. However, for reasons of economy and practicality,

---

[31] There is a long standing debate among qualitative researchers about the relative merits of ethnography when compared to the value of interview data. Becker and Geer (1957) go so far as to argue that participant observation is 'the most complete form of the sociological datum' that provides a benchmark against which to judge other data gathering approaches and to 'know what orders of information escape us when we use other methods'. Discussions of the shortcomings of interviews as a method revolve

especially while researching expert communities that are harder to get sustained access to, interviews are an apt method. There have been very few studies that have managed to get the depth of access required to employ ethnographic methods within government departments. These include the work of Rhodes (2005, 2011) who studied everyday life at three middle ranking British national government departments from 2001 to 2005; and Lewis and Glennerster (1996) who introduce the term ‘administrative anthropology’ to refer to a participant observation-based engagement to understand the implementation of community care reforms in local government. Access in these projects was obtained through former work with these policy communities and negotiations with high level contacts. Despite the severe constraints of the pandemic which made face to face ethnographic work impossible, as the following chapters show, a main strength of my research is the depth of access I achieved after several challenges. Given my positionality as someone who has previously worked with government digital teams, I was able to develop an in-depth, immersive engagement with the government departments; and a sustained rapport with data professionals interviewed in these case studies in order to elicit sympathetic responses using prompts that draw on aspects of the ethnographic approach.

The inadequacy of narrative methods to differentiate between what people say and what they do is another issue flagged by critiques (Deutscher 1973, Hammersley 2012) that takes on added significance in this research. Discussions of the shortcomings of interviews as a method revolve around how it isn’t always possible to know if the respondent is telling the truth (Dean and Whyte 1958, Warren 2002) as one relies on an account of actions that occurred elsewhere in time and space. These issues that stem from relying on narrative methods do not invalidate the usefulness of interviews but highlight limitations that need to be offset by following up on discrepancies and probing for more detail than what is offered on first reply.

This is particularly the case as heightened awareness of the sensitivities of talking about algorithms and predictive models can place limitations on what a mid level civil servant can say to an external researcher (these issues are unpacked in chapter 8). It is often the case that

---

around how it isn’t always possible to know if the respondent is telling the truth (Dean and Whyte 1958, Warren 2002) as one relies on an account of actions that occurred elsewhere in time and space. These issues that stem from relying on narrative methods do not invalidate the usefulness of interviews but highlight limitations that need to be offset by following up on discrepancies and probing for more detail than what is offered on first reply.

various forms of social desirability bias enter the data if respondents represent their work and themselves in an idealised manner. It is possible that when asked about their practices, rather than report routines and norms accurately, they might produce answers that they think are socially acceptable. This might include posturing to signal the ethical use of data technology or other responses that are calibrated to present aspects of their work in a favourable light. This issue was overcome by triangulating with documentary sources and strategies of 'active interviewing' where interviews were not treated as an objective window into truth but as a means to 'get a better understanding of the meaning that interviewees attach to particular events (Kvale 1996: 105)'. I also used elicitation methods and stimuli within interviews where flow charts of previous (paper/discretion based) decision making models are identified from documents produced by government audits[32] and asked directed questions about what has changed in various stages of the workflow (see figure 1 above).

New ways of modelling problematic behavior, adverse childhood experiences and interventions based on them reflects how different kinds of information has now come to be valued; and represents a shift in what is 'knowable' and therefore acted upon (Dencik et al. 2019). I investigate this foundational change activated by datafication through an interview and document based approach by developing two case studies.

Drawing on the conceptual framework proposed in the previous chapter, I demonstrate how welfare provision is being reconstituted through data. In the empirical chapters that follow, Chapters 3 to 4 presents findings from the local government case study and Chapter 6 draws on the central government case study. Chapter 6 focusses on the justificatory discourses (RQ 3) and draws on material from both case studies. The last empirical chapter (Chapter 7) proposes the key theoretical contributions of this project by introducing the key finding of 'reflexive dataism' among data scientists working closely with these technologies.

---

[32] Quality assurance reviews conducted by the central government and produced as open reports have details of the analytical models and business case modelling being used to inform policy in various sectors – published here on the gov.uk content delivery network of the National Assets Register https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/183949/DFE-RR199.pdf

Chapter 3

Datafied 'at risk' categories in Child protection systems

This chapter develops an in-depth case study of a datafied model that has been placed at the heart of a bureaucratic workflow in order to understand how it works in practice; and how the imperatives of datafication interact with existing organisational dynamics at the local government level. Uses of this model are contextually and situationally shaped and the next chapter pays particular attention to the micro level practices co-produced by child protection staff implementing these systems.

> *…And so typically, what you need to be dealing with in a week is a decent, prioritised sense of where your threat, harm risks are… So you want to make sure that you are prioritising the right tactical risks in any given week. You want to make sure that you're giving the right multi agency support to the right children. And at that point, you have to define what the 'right' looks like… what prioritisation looks like. Because there is... given the amount of data that's incumbent on a daily basis, the capacity doesn't exist for there to be an overview that's informed by absolute [pause]… you know, qualitative in depth case based research.* ***So, you have to take a data driven approach***
>
> *Interviewee 14, Business Intelligence Developer, Project Vision C, Aug 2021*

Given the heavy daily caseload, in order to target available resources, the above manager asserts that a data driven approach would prioritise those at high risk. This risk assessment is done through a sophisticated process that produces a composite score based on three different computational techniques. The first method generates what the project designers call an 'index score' after parsing through crime and police intelligence records for mentions of the young person's name. It then compiles a large number of variables based on the frequency, recency and gravity of flags and warnings against their name in key statutory service databases such as school attendance and exclusions, rent arrears, domestic abuse and mental health problems in the family.

The second is a 'network score' meant to indicate peer association risk where the individual is assigned a high score if a social network analysis visualisation tool shows them as being linked to people who already have a high score. The third is a predictive score of the likelihood that a child or young person will fall victim to some form of sexual or criminal exploitation based on a Machine Learning decision tree model that assesses their similarity to previous victims that the council works with through the NGO Barnardo's. There are a suite of such risk models including similar machine learning systems that evaluate whether a school pupil is liable to end up as someone who is Not in Education, Employment or Training (NEET) by regressing against an existing target group.

These metrics are combined to produce a composite score between 50 and 100 that is fed through a dashboard for the use of children's social care professionals at the local council. Sorting these scores into categories creates a priority list that is automated to update every week based on incoming incidents. In general, tier 1 comprises of the names of children and young people (under 18s) at high risk, tier 2 is in the periphery and tier 3 is thought of as emerging risk.

The datafication logic of computationally deriving risk in the above manner is ostensibly intended to prioritise caseload and install clear distinctions between 'tiers' in order to make them administratively legible and manageable. In doing so it enforces a classification structure, it solidifies digital traces of transactions with welfare agencies, propagates an enforcement perspective as it is partially built on police metrics and activates soft surveillance practices. In order to understand how these scores and risk models are transforming the child protection system, the obvious respondents would have been social workers, family case workers and call screeners at the front door to the council who take all referrals from the police, from schools and from members of the public before triaging it and sending it on to the relevant team. These frontline staff members and users of the model were extremely busy safeguarding children during lockdowns and were working overtime on most days as the nature and scope of their roles had expanded during the pandemic. Despite trying several times to secure interviews with them, I had to decide to move from the frontline level to their team leaders for pragmatic reasons. Therefore, in order to build this case study, after the developers and data scientists, I interviewed Education Inclusion Managers, Safer Options team Managers, Principal Social Worker, School safeguarding professionals and youth work providers as users of the model. Most of these managers and team leaders still had input into case conferences, managed 'frontline staff' and had an overview of how the workflow has changed following the introduction of datafied models and the majority of them previously were junior social workers (who were later promoted). Therefore, these respondents had a unique vantage point, had been around for long enough to remember when decision making was based purely on human inputs and brought a rich, historical perspective on how processes within child protection systems were changing over time.

3.1 Category Making

Risk monitoring tells us who we need to be worried about, you can then look at your top 20, top 50, whoever, and then put them into tiers.

Interviewee 8, Programme Support Manager, Early Intervention and Safer Communities, Project Vision C at Local Authority, June 2021

The council's Safer Options team and Violence Reduction Unit meet on a weekly basis to discuss the 'risk cohort' highlighted by the scoring system. Multiagency teams have always met regularly to talk about safeguarding concerns by focussing particularly on different disadvantaged geographic areas in the local authority district. Since the start of the risk modelling process in 2018, however, young people whose names were generated by the system were not previously known to council workers either as part of child protection plans or existing case work engagement through family support programmes[33].

Within an environment inundated with time pressures, the need to prioritise seems common sensical and necessary as the senior developer attached to Project Vision C (Interviewee 14) argues. A profoundly significant change is initiated by choosing to define the 'right person' who needs the attention of child protection workers as the one to whom the model assigns a high score. This in turn creates a ranking hierarchy and arranges the full population of children the council owes a statutory duty of care to on a gradient. Sorting and marking these individuals off into 'tiers' or categories imposes a classificatory logic and this datafied system seems to have unearthed children who were previously unknown to the council. This prioritisation mechanism thus turns out to be much more than just that; as it defines and generatively produces a 'risk cohort' to focus attention on. There is an established body of research within the sociology of statistical methods (Hacking 1982, Hacking 1986, Fourcade and Healy 2013, Beer 2016) and the history of census instrumentation (Desrosières 1998, Bowker and Star 2000, Bouk 2015) that proposes that classificatory labels used by governments do not merely help administrators to read, understand and describe the population they seek to manage. Instead, categories created by large scale data collection systems transform their objects as they describe it and encode the world in particular ways.

---

[33] "So by using that, Qlik app enabled us to start identifying the right young people who we should be concerned about." Interview 17

These tiers are not stable, objective reflections of natural phenomena as it is an ad hoc, 'data driven' way to signify risk. In order to appreciate the power of this kind of 'category making' where individuals are partitioned, ordered and allocated to risk categories, it is necessary to consider the long tradition of research in the administrative context that calls attention to the artificial ways in which categories group and divide people while reproducing a veneer of natural objectivity (Hall et al. 2006, Fischer et al. 2015, Yanow 2015, Yanow 2015, Hjörne and Evaldsson 2016). Categories such as 'tier 1 high risk' and 'emerging risk' acquire a certain facticity when organised in this manner. This implies that there is no internal differentiation between the individuals within the 'tier 1' category. As a corollary, such a definition claims that the attributes of individuals within one tier is different to those in others.

The flattening effects of categories is best illustrated by an example provided by one of the Education Inclusion Managers at the council[34]. He works with the NEET risk model to put interventions in place for Year 11 leavers who are going to be at risk of not accessing learning at the end of the academic year and additionally at risk of being exploited. The system regards a pupil who has been to more than four schools in their academic career as someone to be extremely concerned about – because such a transient lifestyle undermines a young person's sense of belonging and is seen as a contributing factor for being at high risk of dropping out of the education system. He spoke of how the model picked up on anomalies in the data and flagged someone as being at high risk when they had a high number of schools on their record as they were 'manage moved'[35]. A 'managed move' is a practice where a student is made to go to another school for a couple of weeks. They are temporarily put on another school's roll so that they can have a bit of a timeout from their original setting, without necessarily having to formally undergo a school exclusion. As the index score is based on a count methodology, without knowing the context of a student's school trajectory, this would read as an abnormally high number of school moves and be assigned a high risk score.

---

[34] The local authority is obligated to get a 'September guarantee', which is a guarantee of a post 16 place in the further education system.

[35] *How many schools in total they've been to, which is quite an interesting indicator because you can start to see that say, you know… when I rag rate some of these young people, you know, some of these young people have been to upwards of eight schools, you know, in their school career. We generally say that one primary school, one secondary school, and then a post 16 division might be sort of a normal school career. When you start looking at young people above four it starts to get very [worrying]... Interview 15*

Seeing ‘at risk’ as a compilation of data signals and classifying a pupil based on algorithmic fit would mean that someone’s circumstances which might actually be appropriate for no more than a tier 3 risk appears to be more dire than it is. Being identified as a tier 1 risk sets in motion a series of interventions and early offers of help. Therefore, a quirk in the data compiled from various contexts might club someone who is at low risk with another young person with genuine safeguarding concerns; and cause them to be perceived and treated as the same because they are in the same category. Being lumped together in a category causes common features among entities in it to become salient; subsequently leading to a young person being seen as more at risk (by association) than they really are. The institutional category of ‘at risk’ sets artificial cut off points and subsumes the variation in the individual circumstances of all those who fall within those parameters to produce a discrete whole. It is an act of commensuration that creates relationships between individuals within that group and establishes that the attributes of contents of other risk categories are different. In this case, risk is seen not as points in a continuum but as a discontinuous attribute that fits into three disparate ‘tiers’ triggering different responses from social workers and lead professionals.

Without detailed interviews with frontline social workers on how they interpret these scores and names on the tier 1 risk cohort before interacting with families, it is hard to comment on the degree to which it gets used. It is also hard to fully appreciate how much the mathematical tidiness of these models influences their professional judgement without access to the working routines of a sizable sample of frontline staff. However, the systems pulls anyone who has a score between 50 and 100 onto an output dashboard with some predictive text to explain the relevant components of the risk. By tallying with other case management systems, detailed records on the child are linked. So for instance, it links to police reports on whether that young person has had missing episodes and produces an overview of their risk scores over time to generate an overview of their history so that lead professionals can form a view on whether this is rapid escalation behaviour or whether there are other long term things that need to be considered. In addition to interagency groups like the Safer Options team and the Violence Reduction Unit, Early intervention teams, first response teams and the Safeguarding Education team monitor the modelling each week to see ‘who's going up, who's going down, what's changing, who's new?’ (Interviewee 8). These teams are the ones who triage the police inquiries and police safeguarding forms before going into schools to investigate further. The score therefore informs the staff member at the start of the inquiry and helps them gauge the severity of the situation. It helps them prepare for what to expect or ‘get that quick sense of how worried do I need to be, before I start looking into this… and making further inquiries’ (Interviewee 8).

If it is high risk, they might look to see if there is already someone in the council working with them in some way and check if there is an open Social Care episode or if the youth offending team is working with them. If the young person is high on the Child Sexual Exploitation Index, then they would call up the relevant specialised team in the Police to make inquiries along the lines of, ‘I've just had a referral for this child, do you know anything about them?’ (Interviewee 8). They might commission a community mentor, to make sure that young person has mentoring in the voluntary sector. The designers of these datafied models see these scores as a starting point that tries to get you to the staff member who knows most about the case. Within a child protection system in the pre-datafication era, the social worker would sit with a child to make an assessment by a ‘signs of safety’ exercise through a one-on-one relationship based solely on human inputs. In contrast, these scores and subsequent risk tiers and lists of names draw attention to different kinds of information and influence professional judgement in distinctly new ways. While there are gaps in the empirical base of this research which makes it hard to explore and definitely state how these models are being used by case workers, processes of child safeguarding are changing in worrying ways. By allocating children into at risk categories, administrative logics are being slowly reconstituted by datafication.

The problems of this classificatory logic is best understood by the way in which the third part of the composite risk score – the predictive score – is derived based on how similar you are to previous victims of child abuse and exploitation. A data-based assessment of similarity to previous victims, determinations of risk are not based on directly observable ‘offline’ aberrant behaviour but on the likelihood of future abuse. If a young person at 12 fits the ‘measurable type’ description of someone who is a child sexual abuse or a child criminal exploitation victim, based on patterns in data, they are assigned a high risk score.

> *The council has a ‘base cohort’ of young people who have been sexually exploited that they work with along with the NGO Barnardo’s. The model works by regressing this base cohort to the point where they were assigned a social worker at their ‘peak unmitigated risk’. It makes note of what their families look like at that point as well. Based on their similarity to the characteristics of this base cohort, the model generates an assessment of what a child who is at risk of sexual exploitation looks like.*
>
> *Interviewee 8, Programme Support Manager, Early Intervention and Safer Communities, Project Vision C at Local Authority, June 2021*

> *…[We ask] what did that person*[36] *look like when they were 12 rather than 16? So we go back four years to look at them then. And then we say, right, who looks similar to those CSEs*[37] *when they were 12 in our current 12 year olds.*
>
> *Interviewee 10, Former Data Scientist (now works with an analytics consultancy) Project Vision C, June 2021*

These CSE models assess someone as being at risk of abuse based on whether they seem similar to a typical abuse victim or the data template of a current abuse victim on file. Extending Bowker and Star's scholarship, John Cheney Lippold in his book 'We are Data' (2017) introduces the operational concept of 'measurable type' to draw attention to the abstracted manner in which models are built on data patterns and a composite picture of a person seeking government services can be compiled based on partial selections from multiple databases in the datafication era. He defines 'measurable type' as a "data template... used to assign users an identity, an algorithmic identification that compares streams of new data to existing datafied models. Categorical membership is assigned based on algorithmic fit…And fit is subsequently based only on what data is available to be measured, so membership - and identity at large - is based exclusively on data" (2017: 47).

In this case study, the measurable type for a child at risk of sexual exploitation is computed through a decision tree machine learning model trained on about a hundred current victims. This is a very small and tenuous cohort raising serious questions about generalisability. The designers admit that they had to make do with it as it was the most stable cohort they had; where there were clean datasets that showed someone being referred, followed by a triage meeting where they were accepted into the Barnados cohort and assigned a case worker after professionals wrapped around that child had come to an agreement that the abuse allegations were substantiated. The data scientist regressed them back to the date of their referral onto the base cohort at the point of their 'peak unmitigated risk' and based on the underlying structure of the data streams of those 100 target children, the datafied model picked out a measurable type category of a *typical* child at risk of future abuse.

Back in the present, if someone is identified by the model as being at risk, designers are unable to explain what the contributing factors of risk are. Based on practice wisdom, experiential knowledge from the pre-datafication decision making process and social work theory, I was

---

[36] Current documented abuse victim

[37] Child sexual exploitation (CSE) victims

told that there were always key indicators of risk and the model was just a means to identify as early as possible those young people who were 'hitting all of the classic worrying things which are generally present with those young people who are involved in criminality[38]'.

With a machine learning model, it is impossible to pinpoint these 'worrying things' with any level of certainty as the model is essentially measuring similarity to a target cohort. Access to education, number of missing person incidents in the last three months before they got identified, school attendance, property/bathrooms criminal damage and indicators of family dysfunction are all contributory factors. It is however impossible to know for sure. This is a process of working backwards from this base cohort of young people known to be exploited and generalising based on characteristics that are not well defined but nonetheless deemed as correlated with risk by the model. A child's data either fits this measurable type category or not – subsequently, at risk status is assigned based on authentication against this datafied profile. Abuse victims who live off the digital grid and whose journeys can't be expressed as data or translated into a machine-readable template in this neat linear fashion are rendered invisible in this risk model. So would a young person who has been put into care in the area and whose historical records remain with another local authority district. Verification against this measurable type would mean potential victims whose trajectories don't match those of the 100 in the original Barnardo's cohort don't count and wouldn't show up as requiring attention as the modelling is heavily dependent on what data is available to be measured.

3.2. 'Looks like an abuse victim': Group-to-individual problem and the sticking point of classification

Based on a measurable type aggregated from historical records and family characteristics, a high score is generated depending on whether you look like what a very limited base cohort of current victims have done in the past. This involves inferring group-based information, judgements about undesirable behavioural traits from this base cohort and transferring population level risk scores into absolute predictions at the individual level. This group-to-individual problem is a classificatory one and information science scholars have referred to such machine learning procedures as a kind of profiling that violates an individual's informational privacy (Leese 2014, Mittelstadt et al. 2016, Binns 2019) and informational identity (Floridi 2011). Certain patterns or characteristics are identified as producing risk

---

[38] Interview 9

through data processing at the ‘base cohort’ or the population level. Individual young people are then brought under scrutiny when they are tagged as hypothetically having these characteristics by such models; thereby linking their life outcomes with others within a dataset because of the aggregated, collective nature of the risk measure. This aspect of the modelling is optimised not based on judgements about an individual’s unique circumstances; but rather inferences about undesirable behavioural traits are made based on their membership in a group with assumed characteristics.

One expects to be judged based on the merits of your own circumstances rather than be appraised based on what ‘people like you’ have done in the past. This is especially the case in child safeguarding where a vulnerable person requires hands on pastoral and restorative support rather than a depersonalised and proceduralist treatment based on aggregated scores. Being automatically scored and classified as ‘a person likely to be criminally exploited’ or ‘a child at risk of sexual abuse’ leads to pragmatic judgements and early interventions aimed at the individual. However these at risk types are not defined a priori based on the individual’s previous behaviour, household composition or tractable risky characteristics but are ‘supposed to somehow reflect the ‘underlying structure’ of the entities that the data represents’ (Von Luxburg et al. 2012).

This process of clustering individuals into groups links a young person who is *potentially* at risk to a ‘base cohort’ or a statistical grouping that he or she is only probabilistically similar to; thereby properties of the group are transferred and affixed on to the individual. This tension between relating population level characteristics to the individual through a datafied process of abstraction remains even after the predictive score is joined up with the other parts of the composite score. Severe explainability problems (that I elaborate on later) arise when it becomes hard to describe to a frontline practitioner why the model has classified a child thus. Nonetheless interviewees repeated that it was necessary to computationally convene categories in this manner in order to ‘identify early the children who are going to become the people that we are worried about’[39]. Researchers (Redden et al. 2022) underscore the pre-emptive logic of simulating ‘an anticipated future in the present so as to act upon it before it can occur’ and call it the ‘compressed temporality’ of predictive analytics (Andrejevic, Dencik et al 2020: 1531).

---

[39] Interview 8

The overall intention of pre-emptively intervening to prevent something from happening by compiling data-based assessments of similarity to a ‘typical’ abuse victim was revealed by the initiators of the project who see their work as unearthing a ‘needle in a haystack’[40]. Many[41] of the early initiators recounted how shock and outrage at some of the glaring failures in the high-profile child protection cases that their department dealt with shaped the ‘data led’ approach of the risk modelling and scoring process. They referred to two specific cases that ended up with serious case reviews that were highly publicised. One involved cocaine trafficking using vulnerable teenagers while the other was about the systematic sexual abuse of young girls in a care placement order that the police and other frontline professionals were too slow to recognise. Public and media scrutiny of these major cases[42] pointed out that police and council staff had previously come into contact with these victims during the course of their routine operations but dismissed underage sexual activity as consensual and failed to identify the severity of the abuse. There was a sense of malaise and rampant worry that a lot more of these instances of abuse were flying under the radar of lead professionals who didn’t have a grasp of who the key individuals were. They were however convinced that the answers were in the system somewhere. Building a comprehensive data warehouse combining statutory datasets that were held separately by different teams and departments was seen as the way to capture this hidden risk. Mining this ‘haystack’ and letting the data tell them what the ‘typical’ abuse victim looks like or the prototype for someone who is about to be exploited is a process of expecting a picture of the ‘needle’ to emerge inductively without looking for anything specific.

3.3 ‘Data led’ approach to identifying hidden risk: cultural push for organisational change and its paradoxes

‘Data led’ in this context gains a particular valence (Fiore-Silfvast and Neff 2013) and meaning. There is a certain level of confidence in the pure objectivity of the scoring process as reflected in the quotation below.

> *Because you ask a social worker and say, what are the three things that concern you most? And you build a model that finds all of those people, and then they'll go, yeah, I*

---

[40] Interview 8

[41] Several of the early initiators, key designers and data scientists who shaped the architecture of Project Vision C do not all still work on it anymore.

[42] These media investigations exposed the involvement of Somali gangs. Local news articles stressed the fact that these teenage victims were groomed for exploitation by criminal gangs of Afro Caribbean origin by using racially explicit language.

> *know all those people. Of course, because you are the one who told me how to find them sort of thing... that will only tell you about people that you already know about…* ***So that's why we picked the sort of machine learning group that says, well, let us, [pause] let the model decide what the important factors are, rather than us predetermine them.***
>
> *Interviewee 10, Former Data Scientist (now works with an analytics consultancy) Project Vision C, June 2021*

When Project Vision C was initially piloted, design choices meant that insights drawn from the dataset were regaled as better than anything that frontline child protection and social work professionals had to offer. So much so that data scientists had to consciously effect a cultural change in their respective departments by gamifying the uptake of digital processes and look up apps like *Qlik Sense* that were used to parse through the data streams and interact with the outputs. Crime incidents and police intelligence records that are fed into the model were not updated promptly and the Constabulary management had to provide incentives to maintain digital records. On a monthly basis, departments and teams that were at the bottom of a league table in terms of inputting missing persons reports and the other records that brought down the overall data quality felt embarrassed when pitted against others who were able to manage the digital input requirements of the data warehouse.

Disgruntled initiators of the project, especially at the criminal justice, enforcement and policing end of the datafied model complained about a generation gap in comfort levels with the new technology. A concerted effort had to be made to convince those who didn’t see the benefit in filling in minutia in the required digital format within 24 hours of recording the crime on file. New benchmark measures included contacting victims and offenders within seven days and then staying in touch with the victim every twenty-eight days. These time based metrics for the completion of such tasks were initially resisted to the point where at the monthly Constabulary management board meetings, the Chief would pick on the lowest performing department head in front of all other department heads. They were admonished and told that they can’t be at the bottom of the league table again in three months’ time. This created a knock-on effect of optimising and fine tuning the logic of the system where everyone was being encouraged to emulate high performing teams.

There were complaints that mid-career staff members who weren’t too junior but had been with the police force for about five to ten years preferred to walk their neighbourhoods, getting to know people and develop an intuitive approach to identifying hotspots for crime. However given funding pressures, increasing crime rates and the former Johnson government’s commitment to get thousands more police officers on the streets, they had to rationalise and streamline staff resources. In this context, predictive datafied models that can run every day on data that is brand new (updated every eight hours) and simultaneously draw on historical administrative records from the last ten years, proved to be a very attractive proposition in offering to identify hidden risk and 'find things that officers might want to know[43]'. I was told that it was this mid-career cohort that was hardest to train as they were set in their ways.

Meanwhile, senior officers with over twenty years of experience had trouble navigating the new technology but were eager to embrace it. They had been in the force for long enough to see waves of technological change every five to ten years and were reconciled to the fact that 'it was get on board or get left behind' (Interviewee 9*)*. Data scientists in charge of implementing a training programme when Project Vision C was first introduced claimed that young newer recruits were digital natives and immersed in the world of visual data, social media and well versed with running analytics on apps. They bemoaned that during the course of the training it became evident that older cohorts of staff who didn't grow up on social media could not navigate through the Qlik Sense app or perform basic tasks like get to the next page or follow a link.

> *If you have ever used any social media application or diving through… [pause] it's very kind of, I don't know, some things I always feel like… [with frustration and exaggerated emotion] How do you NOT know how to get to the next page or follow a link?*
>
> *Interviewee 9, Integrated Analytics Hub Manager for Police Constabulary, July 2021*

Once this self-reinforcing rhetoric of needing disruptive organisational change in order to find the 'needle in a haystack' was in place there was no backing away from it. Being desperate to prevent another major catastrophic child protection failure, these risk models were thought of as just another decision support tool that helped raise awareness around signs of exploitation. In seeking to pre-emptively identify signs of exploitation, neglect and abuse by looking through a 'haystack' in this manner, these data signatures of 'people who look like' past CSE/CCE victims are solidified and encoded into the system.

This is worth highlighting as the variables that go into these risk models are socially constructed and not pre-analytic, objective reflections of reality as Interviewee 10 seems to think when

---

[43] Interview 9

arguing for a ‘data led’ approach that overrides the preferences and instincts of social workers. The composite metric that is finally produced from the index, predictive and network score relies on transactional data generated by the digital imprints that recipients of council services leave behind during routine tasks such as paying rent for social housing or sending children in the household to the local state school[44]. The simple act of sending children to the local school causes one to lose control of their digital footprint as the pupil can be tracked and identified using unique learner numbers that are meant to lapse at the age of sixteen but tend to follow them throughout their school career as there is very little clarity on what should be preserved, collected, and retained. These identifiers are often widely distributed to service providers such as youth support work organisations and even to the central government, repurposed and linked with other data such as their Child in Need status – private school students are significantly less visible and tractable. Internal information systems of the council also capture digital traces when someone accesses help for domestic abuse or mental health support; making the collation of these datasets in order to draw inferences about the salient characteristics of those with high levels of need a sensitive endeavour.

This is especially the case for the datafied model being used by this council as they take datasets from the justice and policing systems which were originally collected to gather intelligence about the occurrence of crime; and combine them with statutory social datasets held by the

---

[44] Refer to the advocacy organisation Defend Digital Me’s The State of Data 2020 report (Defend Digital Me 2020) for extensive analysis of how at the age of 5 children lose control of their digital footprint just by virtue of going to school. In addition to school attendance data, information on the names and addresses of pupils and their parents are collected through the school census. Attainment data from Early Years to A-levels tests, school exclusion lists and details of those eligible for free school meals are captured. Each student in the state school system in England has a unique pupil number (a 13-character code that identifies each pupil which is expected to remain with them throughout their school career regardless of any change in school or local authority). This unique pupil number (UPN) is linked to Child in Need status of under 18s that the council have a statutory duty of care for and there is very little public information on how consistently this is managed even though the UPN must lapse when pupils leave state funded schooling, at the age of sixteen or older. However, by age sixteen it has been widely distributed, and linked with other data including the unique learner numbers which take over identification post-16 as there is very little clarity on what private information service providers such as councils, youth support work organisations and others should collect, preserve, and retain. It is worth noting that children who go to private schools have more control over whether their educational pathways are this visible and tractable to central and local government bodies.

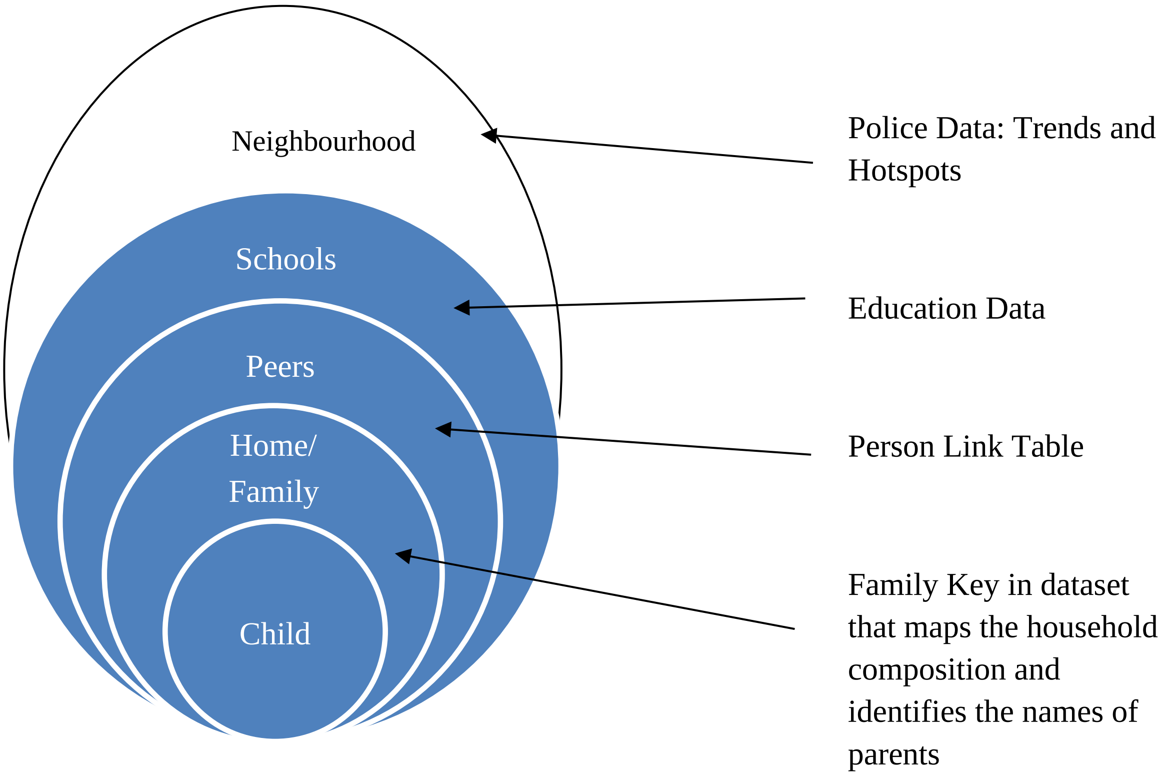
Furthermore, repurposing of educational records has become normalised at the central government level. Since July 2015, the Department for Education has facilitated the Home Office’s monthly matching of children’s national pupil records with Home Office records to find people for immigration purposes under a Memorandum of Understanding on data sharing. This data linking leading to detrimental context collapse is in progress without the Department for Education demonstrating any accountability or concern for what happens to the families as a result of this exercise.

council. The appropriateness of repurposing, combining and building on intelligence products about Anti-Social Behaviour and large-scale police investigations into child sexual exploitation (like county lines drug busting operations) is not questioned. During the early stages of the project, neither was it ever asked if confidence in the objectivity of these biased, flawed datasets is warranted.

In the initial stages of the piloting of the project, a key professional competency that was missing was the duty of candour in considering the adverse impacts of linking and combining fragments of information about a child and family held in different databases (see Figure 3). As the project progressed, it was only after two years into its implementation and embedding into institutional working routines that frontline professionals and project initiators discussed whether the data points were actually indicating what they think they were in terms of foreshadowing abuse, vulnerability, and exploitation. There is now a growing field of research that documents technology augmented abuse of power by the police (Murray et al. 2020, Fussey and Sandhu 2022, Oswald 2022) on a systemic scale where criminalised individuals are surveilled in order to manage 'deviant' populations and nudge them back into normalcy. Function creep where these datafied tools are used beyond their initially specified purpose is mostly invisible as the law protects and is in part shaped by the police. Advocacy organisations have pointed out a locker room culture in British policing that is rife with misogyny. A landmark super complaint (Centre for Women's Justice 2020) submitted by the Centre for Women's Justice offers evidence of the misuse of law enforcement databases to perpetuate intimate partner violence; and a culture of impunity where known abusers within the police force were using their knowledge and access to police systems to shift the blame to the victim (see the joint response from the College of Policing, Inspectorate of Constabulary and Fire & Rescue Services and the Independent Office for Police Conduct that concedes to these points - College of Policing 2022). There has also consistently been evidence of higher rates of domestic violence amongst serving members of the police workforce when compared to the general public, accountability problems and cumulative reports of blatant violations of police procedure that are hard to ignore.

In this light, putting systems in place to identify hidden risk and letting police officers 'find things that they might want to know' (Interviewee 9) sets in motion a self-reinforcing logic where criminalised communities are further disenfranchised. The abstraction and formalisation of the prototype for an 'at risk' category not only defines normalcy but has a secondary function to enforce it. These structuring and generative effects mean that by defining risk and normalcy through typification and data prototypes, these processes have the power to prescribe how things ‘should’ be as well as provide concrete measures of deviance from the norm (Canguilhem 1989: 237).

Figure 3: The ‘Haystack’: heuristic of the datasets that are combined in the Data Warehouse on which the Risk Models discussed in this chapter are run



This is a cause for some concern as the data warehouse captures information about the family of the child or young person in question by developing indicators of whether someone linked to the child in the family is a victim or perpetrator of domestic violence or has been cautioned for non-crime domestic violence. Policing cultures do not have a good track record of managing and responding to domestic violence (DV) issues in a restorative manner or in adopting a stance that is empowering to the victim-survivor. There are countless accounts of untrained police officers confusing the perpetrator for the victim, misrecognising the power relationship in instances of bidirectional violence and consequently making the abusive situation worse for the victim. Furthermore, while domestic violence call outs might have police records that can be counted as separate episodes and cumulatively scored to form a view of the household environment, things like emotional and psychological abuse do not have any ostensible trace.

If the aim is to look for indicators of family breakdown as impeding the child from flourishing, this exercise defeats the purpose.

Policing databases have a material infrastructure and this process of datafication unmoors records that are generated by a frontline cop back at his desk after a domestic dispute call out while filling in a routine digital report. He would have no idea about the external audience for this file let alone a grasp of the centralised risk models that will parse, collate and aggregate that record in order to create a risk cohort. Therefore this 'data-led' orientation marks a key shift away from the indexical, documentary and evidentiary function of the record (Currie and Hsu 2019) towards its reuse for other purposes such as to compute whether this isolated incident will lead to a pattern of abuse, exploitation or a welfare concern for the dependent child linked to the household.

If families at the receiving end of statutory services become aware of the fact that a certain number of domestic violence (DV) marks against their household would lead to interventions into the life of their child; or a risk score that flags them for the focussed attention of inter-agency support professionals, they would be hesitant to ask for help at all. This can potentially lead to a scenario where women in domestic violence situations suffer in silence in fear of having their children removed if they report it to the delegated statutory contacts. Early findings (Edwards et al. 2022) from a large ESRC research project investigating parents' trust in the operational use of data linkage and predictive analytics identifies an overarching rationale of solutionism which the authors refer to as 'problem-solving for the sake of problem solving'. After conducting a nationwide survey of 843 respondents from across the UK[45], semi structured interviews and a comprehensive document analysis Edwards, Gillies et al find that the social license for building service interventions for families by linking data held on them and their children is very low.

One of the foundational theoretical interventions of Critical Data Studies is that data is necessarily othering. Discussions about data in this body of literature come from scholars occupying a broad spectrum of disciplinary positions from Anthropology to Policy and Legal

---

[45] The point of the survey was to see if parents agree or disagree about what is acceptable or unacceptable in relation to linking data collected by government services and using predictive analytics for risk modelling and intervention in families' lives. This project notes that while such ambitious data linking is encouraged by the National Data Strategy (DCMS 2020), the enthusiasm for problem solving using predictive analytics defines the problem as a purely technocratic one and diverts attention away from its complex structural causes.

Studies and Digital Humanities. Science and Technology Studies scholars writing within this field (Gitelman 2013, Kitchin and Lauriault 2014, Dourish 2016) argue that data must not be taken as a given as it is produced through material practices and logics of equivalence that translate the phenomena it is meant to represent into observable reality. Data is always constructed and there is never a one-to-one equivalence between objective reality and systems of quantification. This especially true when records with a distinct function and audience within the statutory care domain and the policing sector are stripped of its original context and recombined, moved, and bundled together. As we have seen, the justification for the decision tree machine learning model offered by interviewees 8 and 10 draws on the metaphor of finding a needle in a haystack in view of the undefined yet imminent threat of catastrophic child exploitation and abuse. In insisting for cultural change in the organisation to push through a 'data led' approach to generating a list of names of children to be worried about, the project proponents think of the administrative data that underlies the training dataset for the decision tree ML model as representative, infallible and highly accurate.

Case files and administrative records, even when created by the most caring social workers, can at best provide a partial view of the lives of children and young people who are the service users. Practitioners and social policy academics have pointed out how even this partial view may be undermined by incompleteness of the record and delays in closing a case as needing no further action (NFA). Managers are not fully aware of shorthand practices where the frontline social care staff fudge their risk rating during their assessment so that families are classified as at high enough risk to be given the services they want them to have (Lyle and Graham 2000). There have also been reports of improvisational tactics where social workers making an assessment would hold a case open for review (in order to devote sufficient attention to it) but log it as complete on the system as they know that time-based targets are being used to evaluate their work.

With the use of electronic case management systems, warnings have been sounded against relying on administrative data to represent the domain of child maltreatment and the extent to which data signatures can be faithfully interpreted as risky behaviour (Gillingham 2019). This is particularly the case with the records held on statutory systems by the council - the architecture of which comes from the Poor Law system from the 18th century - but is still overwhelmingly dealing with low-income families, ethnic minorities, single mothers and those on other forms of heavily stigmatised assistance. Adding policing and justice datasets to this bureaucratic classificatory architecture and then training predictive models as decision support tools creates a potent mixture. The ease with which the rhetoric of ‘data-led’ is used thoughtlessly, carelessly and interchangeably to justify a datafied system whose sole purpose is to trigger awareness of where the future child exploitation case load is going to come from is therefore disturbing.

In attempting to trigger awareness in this manner, the system as a whole becomes intrusive and increasingly targets and redlines marginalised groups without intent. Awareness of where case load is going to come from might be useful from a practitioner’s perspective but system usage data is not a good indicator of child need. What this measures is the episodic nature of vulnerable people’s engagement with the council. Tracking case volumes and metrics that count the number of times someone interacts with the council or accesses help is unsuited to wholistic ongoing care that is meant to support a young person through a crisis. It also conflates statutory social care and the principles of early intervention by blurring the threshold of significant harm where frontline professionals are supposed to make concrete determinations about the welfare of a child and subsequently make contact with them. In other words, the risk scores and tiers that have been introduced by the datafied model do not correspond to the formal legal threshold for removing a child from their family or the Local Authority’s duty to investigate (section 47) or the responsibility owed to Children in Need (section 17) in the area. The scores, rather than reflect objective reality, are a shorthand for risk that might just be high because the family in question has a lot of data on record on them through being in arrears with their social housing rent, or having been in the care system or through other routine transactions with the council while applying for income support or for help with additional complex needs.

Other indicators that contribute to the score that each child is given are concern for their mental health, teenage pregnancy and any involvement of parents with the police or courts and probation. It was never precisely explained to me how the signs of concern for the mental health of a young person translates into these risk categories described above. Various promising protective factors at the intersection of mental and physical wellbeing have been identified by social psychologists. This means that despite being exposed to adversity and trauma in childhood, there are various protective factors that can mitigate Adverse Childhood Experiences (ACEs) from turning into negative health outcomes across the lifespan. Banyard et al. (2017) posit that there are aspects of the individual, family, and community that can promote resilience and good health despite exposure to adversity, childhood victimisation and financial strain. There is now an emerging field of research that advocates for a strengths based approach to child maltreatment prevention based on the hypothesis that a range of strengths across the competency domains such as emotion regulation, meaning making, community support and social support can lead a child to go on to thrive and enjoy a good quality of life despite facing significant adversity. These protective factors that contribute to a wellbeing (rather than a risk) orientation are sociological in origin and are harder to capture neatly into a data template that can be accommodated by this system. Therefore, mitigating factors such as, say extra familiar support for a child coping with their parents' divorce and subsequently having mental health issues, are not considered and negative experiences cumulatively pushes someone's risk score higher. This is significant; as seeing someone as a product of Adverse Childhood Experiences (ACE) counts alone using these novel data technologies that build actuarial instruments and measures for risk represents a profound normative shift that reconfigures the social care professional's capacity to decide and act.

Datasets held in the data warehouse are structured by institutional categories and historical systems of social exclusion. The epistemological positioning of the project rationale[46] where 'data led' meant drawing on data in an apolitical, 'pre-factual, neutral, objective and pre-analytic' (Kitchin 2014) manner leads to the naturalisation of the subsequent 'at risk' measurable types that emerge from the machine learning 'haystack'. This is a process of translation where there is no appreciation for the tacit institutional knowledge that determines how records are organised, formatted, or tabulated. These data technologies are embedded into organisational cultures where insidious political imperatives determine whether a domestic violence (DV) episode or an Anti-Social Behaviour injunction[47] or a breached order is even recorded in policing databases. Anti-social behaviour is such a subjective construct that it can mean anything from noise complaints to teenagers hanging around on the street indulging in boisterous social activity that can create a sense of anxiety and menace in the onlooker. Policing by consent is a principle (and not a duty) that cops don't always uphold when enforcing limits on public gatherings and other law enforcement functions that could lead to a young person

---

[46] This characterisation is especially true of the early version of the project rationale as its objectives and operationalisation has evolved over time.

[47] The Anti-social, Crime and Policing Act 2014 broadened the definition of ASB and made these powers easier to use and available to more agencies. New injunctions to prevent nuisance and annoyance and Criminal Behaviour Orders can impose positive requirements upon individuals as well as prohibitions.

getting a charged or having their name taken down in an incident report. Seeing someone as a product of their Adverse Childhood Experiences (ACE) count aggregated from statutory council databases and overlaid with policing intelligence risk models derived from these kinds of value laden datasets is incredibly reductive. Crucially, it can prove to be a counter-productive start to a relationship between a social worker with a family they are seeking to introduce themselves to before putting early offers of help or interventions in place.

Chapter 4

The ‘needle in the haystack’: from risk-based prioritisation to peer association modelling in Child Protection

During the course of this research, I conducted interviews with data workers and those close to the implementation of these datafied models in order to document day-to-day practices and justificatory discourses. While a detailed investigation of service user perceptions and the impact on the thousands of dependent children and vulnerable young adults is outside the scope of this project, it is important to hear their voices so as to appreciate how disempowering reliance on formal case files and a social worker approaching a vulnerable young person based on an overreliance on the ‘paper self’ can be. The below quotations are from the empirical material of a large study exploring the over-representation of care-experienced girls and women in the youth and criminal justice system across England but focussing on a pre-datafied case management workflow (Fitzpatrick 2020, Fitzpatrick et al. 2022).

> *“I felt like the social workers that had worked with me just read my file and formed an opinion on me without actually knowing me. They wasn’t always around, and their solution is always to move me or blame me. So no, when I was younger, I never felt supported”. (Bobbi, 20)*
>
> *“My support worker at the care home, one of the things she said that really peed me off was that she was like ‘I’ve read your file, I know everything about you’. I was like ‘no, you’ve read my file, you know what’s on my file, I am more than just a piece of paper’”. (Ellie, 18)*

Guided by insights from feminist criminology[48] and a focus on lived experience (Burman and Gelsthorpe 2017) this research underscores the negative gendered and racialised judgements at different stages of the youth and criminal justice systems that lead to challenging behaviour of girls in care to be unnecessarily criminalised. There are overlapping layers of disadvantage for girls from black and minoritised communities as represented by the quotes above. Fitzpatrick (2022) presents irrefutable evidence that gender stereotyping of young girls in the care system draws on patriarchal constructions of what ‘ideal’, ‘innocent’ and ‘deserving’ victims look like. She demonstrates how professionals’ gendered judgements of girls from ostensibly ‘chaotic’ backgrounds cause them to be less likely to be seen as genuine victims. Girls at risk of child sexual exploitation are often described as teenagers out of control,

---

[48] This is a Nuffield Foundation funded project that rests on the premise that statutory care provision needs to be so much more than just providing accommodation.

manipulative and difficult to engage. Seeing them as sexually knowledgeable and/or sexually active removes them from the category of the 'child' that needs protection as they are no longer innocent and from the category of the 'victim' of exploitation as their innocence is no longer in danger (Woodiwiss 2018). Perceptions are clearly important in guiding professional response and action. While the victimisation and exploitation of girls in the care and youth justice systems is often overlooked or minimised, participants in the study described how police call-outs to care homes occurred for behaviour of girls that would not necessarily result in police involvement for other children. Looked after children are more likely to be criminalised and girls in particular felt their behaviour was 'under a microscope' due to their care status.

Data tools that are not looking for something specific but using properties of data points to find out what one is looking for can reinforce negative judgements of those in the care, youth and criminal justice system and reinscribe these abuses of power and misogyny. When stripped of the rhetoric, this is a process of searching for 'the needle' or a data signature of a potential CSE or CCE victim based on culling through the haystack of routinely collected administrative data to find young people who look like what current victims looked like a few years ago. Forecasting child abuse and identifying potential and emerging victims of sexual and criminal exploitation depends on the properties of the starting dataset of welfare histories and the records of those currently involved with statutory agencies. The above discussion has clearly shown how those from disadvantaged backgrounds are disproportionately overrepresented. This could be because they are more likely to be already on the system whether through misogynistic unconscious biases of professionals criminalising problematic behaviour into these official records; or because low income families on stigmatised forms of assistance are more likely to have extensive previous interactions with welfare services and thus have their child scored as being at high risk simply because the system has more data on them. The incompleteness of many of the datasets that are being used for the modelling is also a concern since child protection datasets are known to be incomplete in a non-random way (Munro 2019).

This raises questions about the appropriateness of these datafied models and pre-emptive risk modelling based on data traces. More importantly it calls into question the confidence in 'data' as a source of authoritative interventions. Anyone on a risk cohort list would be more likely to be stopped by police or reached out to by youth offending teams (YOT) or social work professionals at the council. However, every time they are stopped or engaged with, they get another data point on the system making it more likely that they are considered as a chronic high risk person and more likely to be open to intrusive interventions in the future. Predictive risk scoring therefore replicates and amplifies the problems posed by human driven social care and policing practices.

4.1 Credibility Practices

The case for populating a list of ‘at risk’ children based on a machine learning model rather than use social worker’s expertise to code in a certain number of school absences, exclusion notices or care status as signifying risk, was made by deferring to the power of data driven analytics processes. Not letting the data lead would be “predetermining what the important factors are” (Interviewee 10) and relying on social worker’s intuitive and practice wisdom based sense of who they should be concerned about. The data scientist while describing the early thinking behind the Insight C project said that if you build a model that finds all of those young people who correspond to the risk factors that are at the forefront of a social worker’s mind, there would be no value addition. He hypothesised that lead professionals would then say “yeah, I know all those people” - to which he would have to reply at every juncture, “Of course, because you are the one who told me how to find them!” (Interviewee 10).

While piloting the system, data scientists tasked with operationalising the vision of the project had to spend a considerable amount of time convincing those interacting with the outputs from the models that they were trustworthy. Various ‘credibility practices’ (Kolkman et al. 2016, Kolkman 2022) had to be activated to establish and ascertain the list of high risk young people that is created by the datafied process. Data scientists complained that when it was first introduced, old fashioned social workers and safeguarding practitioners would not trust the output of the model if they don’t see names that they are used to seeing in the list of vulnerability. In this case, data scientists tended to slip in some of these known names in order to foster trust in the model.

> *So, with our missing children, we had to... So we had a top missing person, a missing child in [a city in England], they were about 14, they would go missing five times a week, if not more. And we built our missing children model and they weren’t on the list.*
>
> *…And in short, they would always go missing at 10 o'clock at night, because the police told his family that if the child is missing at 10 o'clock ring the police, but [his mum] would let them go out and meet their friends at nine o'clock. As soon as it hits 10, if the child wasn't back, she'd ring the police and go ‘They are missing’. The local officer would go, ‘well, I know they are hanging out… at the co-op car park’. He would literally turn up and say, ‘go home you, your mum's called us in’.*

Interviewee 9, Integrated Analytics Hub Manager for Police Constabulary, July 2021

The intelligence officer in charge of implementing these risk models after the founding data scientists had moved on to better paying jobs at specialised analytics consultancies in the private sector recounted how by the time policemen had reached the spot where the allegedly missing child was known to frequent, this young person would have gone home by then. Almost every member of staff in the police constabulary knew this 14 year old by name as they had taken them home on a squad car at least once during an overnight shift.

The lists created by the datafied missing children model however did not have their name on it. This led to a serious contradiction where safeguarding and policing professionals would not trust the output created by assiduous data scientists because the one, most frequently missing child they had all met and knew by name was not on the list. So the key dilemma that the designers of the project had to confront was how to provide data that is interesting and new so as to identify 'hidden risks' while at the same time convincing people that the list was working by telling professionals what they already know. Slipping in the name of the 14 year old, whose missing person incident reports are innocuous instances of harmless frivolity, is a clear act of not being 'data led'; but leveraging the practice wisdom and expectations of frontline staff to validate the new datafied process. Even while claiming to be led purely by data in identifying those at risk who wouldn't be picked up by the pre-datafication process, in order to foster trust, initiators of the system have to strategically construct certainty in the machine learning model. This is achieved by deemphasising the accuracy of the previous ways of working through calling into question the expertise and situational judgment of safeguarding professionals. Calls to unearth hidden risk signatures of abuse victims through data imbues the output from the machine learning model with epistemological authority; whilst making uncertain the intuitive reasoning (Munro et al. 2017) of cohorts of frontline policing and social work staff who are set in their ways and need inducement to change.

Even while activating these credibility practices, members of the data analytics team admit that there are others who go missing not as frequently as the above 14-year-old but when they do, they were found a few days later in 'less than nice circumstances'. These were the genuinely vulnerable young people that they wanted and needed to protect. However, the name of the above teenager needed to be on the list to convince people the list was working. The authoritativeness of the data and the scoring that relies on it had to be constructed and these contrivances were necessary to validate the claim that there is a hidden risk out there to young people's safety that only the datafied model can unearth[49]. This example also shows the sometimes problematic logic of equivalence where relying on the number of times someone goes missing or the total amount of time they are listed as missing as contributing to their high risk status may not capture the seriousness of the situation.

By structuring the instruments of risk on data on service usage and the measurable type of a past victim, prioritisation of casework is drastically reorganised. In doing so it exacerbates the professional practice in the child protection system that has over the last twenty years moved towards managerialism, culpability and objectivism. The social work profession, especially with regard to safeguarding functions within local government, has been slowly changing into one organised around where to lay blame (Griffiths 2000). The mandate to prevent child death and abuse, often in the face of sensationalised media coverage, is such an emotional argument. This means that despite practitioners and academics sounding the alarm, culpability rather than a processual orientation that values a child centred mode of working (with the voice of the child being systematically heard) has become the norm.

I heard from the Principal Social Worker[50] in charge of managing around 800 frontline social work professionals that as of now his is a profession like no other where one mistake could end his entire career. By referring to how they couldn't afford to make any mistakes as even a small one would result in the death of a baby that would then be politicised and publicised, he spoke about how individual social workers tended to be 'subjective' in their assessments making it a managerial problem. As a consequence, attempts to institute 'modern' grids and checklists that standardise their work continued apace even before the datafied systems were initiated to combat potential life-threatening mistakes by putting pressure on junior social workers.

In such an environment, there is an overarching preoccupation with risk. This is manifest in a managerialist impulse to set systems in place to continuously predict need in the local authority;

---

[49] This is in keeping with findings from the Sociology of testing of new technology and Critical Data Studies literature (Benchmann and Bowker 2019, Amoore 2020, Grill 2022) which shows how claims of certainty in Machine Learning was accompanied by efforts to establish that it is working well. Attempts to establish credibility in this manner almost always involves political choices that usually seek to validate that the model is good enough to be adopted and scaled up by discrediting previous modes of working.

[50] This respondent is based at another council where I conducted preliminary exploratory fieldwork through informal interviews. These sentiments were however shared widely and several interviewees repeated how it would be careless to not want to prevent the deaths of children like 'Victoria Climbié' or 'Peter Connelly'.

so that support workers can intervene earlier and refer clients to other services to help prevent a crisis situation from worsening. This overarching obsession with risk goes as far as ambitions to identify potentially recurring cases at the point of assessment. During various interactions with frontline users of the full datafied model in use at this council, it became evident that there is a way to toggle through the data warehouse's user interface to reach an individual young person's 'vulnerability view'. On this webpage, safeguarding professionals such as education inclusion managers and social workers can see all that is going on with that young person in terms of what has happened in the last month. At the time of the interviews there were plans to map various incidents that are of interest to the data scientist team against school attendance data – when graphed over time this displayed the full pattern of school exclusion, involvement in criminality and the concomitant dip in attendance. There were ambitions to develop further predictive models in order to use data to make statements like "whenever this happens, it equates to an average of X percent decrease in attendance and therefore if this certain combination of factors happens, then typically, this means that 'this then happens afterwards'" (Interview 15).

There is a causal certainty in the language used to attempt to pin point the linear directionality of a drop in school attendance leading to an adolescent showing up on an criminal incident report or being tagged on police intelligence systems. The school attendance of a child or a young person could have dropped for any number of reasons. This line of reasoning obfuscates the fact that child maltreatment in households is notoriously hard to identify. It becomes hard to assess if the people who are being flagged as being 'at risk' by these datafied models are actually at risk or just an artefact from these flags being derived from datasets on service usage. This enthusiasm for putting predictive percentages on everything in order to prevent an undefined yet imminent harm draws on an 'objectivist' and causal understanding of family violence which see one shot altercations between an aggressor and a victim in an isolated manner. It assumes that one person (A) can cause another (B) to do something - for instance a domestic violence incident that was provoked by the female partner dressing in a certain way (Greenland 1987, Griffiths 2000, Oak 2016). The inability to see the interacting variables in 'high risk' families and the non-linear, ongoing nature of domestic abuse gets exacerbated when risk metrics and quantified scores are applied to complex phenomenon. As is evident from the discussion on the social embeddedness of the case management electronic database held by police and statutory teams in the local council, what looks like a single incident of a police call out where partner A was being violent towards partner B could be part of an ongoing series of often invisible abuses of power in a relationship where the abused partner could have resorted to violence to fight back. When these historical DV incident reports are repurposed, collated, and aggregated into an alert on the dependent child attached to the household address, it is hard for the frontline social work professional designing interventions of support or early help to disentangle the causal factors to family breakdown. Furthermore, it is hard to identify who the abusive partner in the household is with any level of confidence without reflexively engaging with the single point quantitative risk score.

Figure 4: Full spectrum of services provided by this council (Universal, Targeted and Specialist) and the thresholds for escalating support needs to social care.

Thresholds

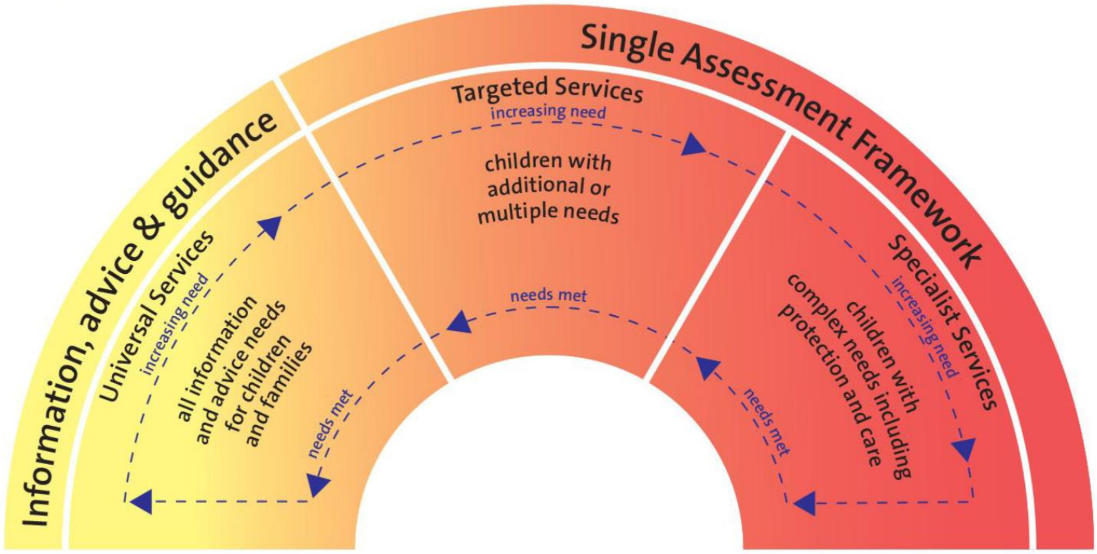


Fig 1 – Levels of need and support

Note: The guidance document conceptualises these needs on a continuum (see arrows) and has protocols in place to step up the case and to step them down as appropriate

This remains the overarching paradox of pre-emptive risk labelling based on data traces that sees digital evidence of problematic behaviour as foreshadowing child abuse and neglect. If we use complex data technologies to predict where those at high risk of child abuse and maltreatment were, before it happened, it is possible to theoretically reduce the risk. However, since what is being assessed is a risk, are these models pre-empting and labelling households who haven’t actually experience these outcomes? By targeting certain types of children (and their siblings in the same household?) who resemble the data signatures of previous victims, is this affecting an outcome that would otherwise have not occurred? The arrival of paper checklists in order to modernise the social work profession twenty years ago was criticised for having a deficit model that aimed to make a call on whether a person has hit a criterion or a threshold for intervention. Prior to the introduction of datafied models in 2018, this council also had detailed guidance in place for a standardised Single Assessment Framework where safeguarding assessments are to be made on the basis of prescriptive, predefined grids and outcomes driven care plan documents. These have been shown to propagate a counter productive focus where the social worker sits down with a vulnerable person with an assessment document and proceeds to have a series of sensitive conversations about what they can't do. Such a start to the partnership does little to generate a bigger picture of the person in terms of their strengths and makes them excessively nervous and can possibly retraumatise them if conversations about abuse and sexual violence is handled incorrectly.

Using metrics and machine learning models to identify potential and emerging risk cohorts takes this managerialist objectivism and risk orientation to another level as there are now more precise ways to target problematic behaviour through digital proxies[51]. Scholars of algorithmic regulation and information ethics (Beaussier et al. 2016, Yeung 2017, Yeung and Lodge 2019) have warned of the dangers of 'risk based prioritisation' (Yeung 2018: 511) where in trying to rationalise and manage the limits of what public sector regulatory standards can practically achieve, the focus shifts to highest priority risks. This means that public sector organisations cannot or should not even try to address all harms and the known issues of disadvantage and deprivation in the local authority. There is now a growing body of statistical evidence on how poverty harms children (Alston 2019) and definitive figures on how many children are in poverty. It is also well known that bad schooling and an underfunded health service has an ongoing impact on the life chances of children and adolescents. In developing diagnostic

---

[51] Legal Gateways for capturing and processing these datasets are:

1. The Digital Economy Act 2017 and the Public Service Delivery, Fraud and Debt: Data Sharing Code of Practice
2. Sections 10 and 11 of The Children's Act 2004,
3. Section 82 of The National Health Service Act 2006
4. Section 1 of The Childcare Act 2006
5. Section 23 and 25 of The Children's and Families Act 2014
6. Section 17, 37 and 115 of The Crime and Disorder Act 1998
7. Section 11, 21, 157 and 175 of The Education Act 2002
8. Section 1 of The Localism Act 2011
9. Section 1 of The Children (Leaving Care) Act 2000
10. Special Education Needs and Disability Regulations 2014
11. Immigration and Asylum Act 1999
12. Education and Skills Act 2008 (ESA 2008)

technical tools to try to reduce the several million children in poverty down to identify the 500 or so ones who need particular attention, datafication corresponds to this logic of risk based prioritisation.

Through formal assessments of the probability and consequences of high priority risks it is hoped that safeguarding professionals can strategically target their services to the riskiest cases and thereby prevent many of the adverse events they worry about. However, in using data led modelling of routinely collected administrative data to predict the likelihood of future adverse outcomes, families from low socioeconomic backgrounds are oversampled as the predictive model outputs depend entirely on the starting properties of the training dataset. In labelling someone as ‘risky’ not only is it stigmatising, it also affects the development of the child and their life chances. In identifying a poor family as needing professional statutory help and labelling them as risky, this stereotype stays with them. The stigma attached to interventions from social workers and lead professionals at the council is so great that most low-income families still see them as being there to take their children away. Unlike modelling from the natural and life sciences where diseases progress in an objectively linear trajectory; criminality, child abuse, sexual exploitation and other social problems do not have a predictable causal pathway that can be gleaned from the data signatures of those who were previous victims. Even if all known historical statutory and policing datasets are crunched and properties of the data are used to find what one is looking for, this predictive scoring is a managerial, productivity boosting tool that prioritises case load efficiently rather than an objective reflection of a prototype for signs of child sexual exploitation.

Therefore, by letting the datafied model decide what the important contributing factors to child sexual exploitation, criminal exploitation or being at risk of dropping out of the formal education system are rather than predetermining them; and by rationalising a ‘data led’ approach as objective, the project proponents are making an insidious policy choice. Letting the data decide which the most risky cases are and then flagging them for prioritised attention has clear disadvantages that are lost under the veneer of scientific scoring and classification. In order to elaborate this further, I discuss two key problematic input variables from the data warehouse that goes into the final score for each child before unpacking the network component of the score in the upcoming section (Section 4.2).

Firstly, in addition to usual tags such as ‘Looked After Child, ‘Child In Need’, ‘Child Protection Plan’, ‘open to Families in Focus’ marked against the relevant households, this council captured information that indicates concern for the mental health of the child or family member (see full list of indicators developed for the risk modelling in Figure 5). In seeking to manage the risks of mental health by collating fragments of data held on the internal information systems of both the council and police constabulary, it is likely that the machine learning model is ingesting frequency of support service usage. Being sectioned under the Mental Health Act 1983[52] would leave formal records if the subject of the sectioning order cooperates or voluntarily goes into the psychiatric assessment unit with the police or ambulatory services. The point of statutory sectioning is to clinically assess if the person is a danger to themselves or others. However there are massive variations in mental health needs. There are also a large number of instances where reports of concern for someone’s mental health is weaponised by perpetrators of family violence and domestic abuse to control, harm, and intimidate their victims. There are few mechanisms in place to check if reports of poor mental health are a data artefact from extensive previous interactions with statutory services, or self reported requests for help with mental health issues or something that has been tagged on by say a teacher in the child’s school who was concerned. The data on mental health fed into the modelling is so woolly and ill-defined that it is hard to distinguish between a mental illness diagnosis by a consultant Psychiatrist and a note on mental health concern made by the Social Worker. Minor and major problems are subsumed under the same heading where concerns for mental health and wellbeing can’t be differentiated from post-natal depression or a psychosis episode – all of which have a massively different impact on the individual.

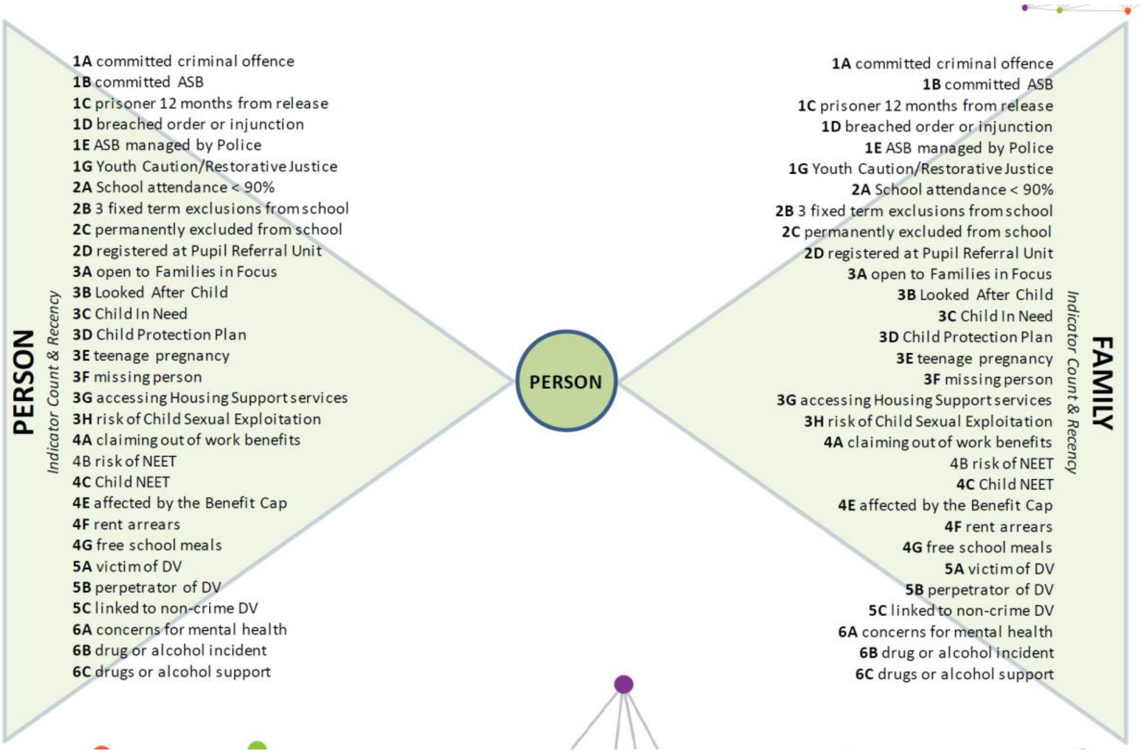
Mental health statutory ‘crisis’ services in the city or local authority run in concert with the NHS are often overloaded and are aimed at making assessments of whether someone is suicidal at a certain point in time. They are less equipped to reintegrate someone back into society or provide wholistic care and long-term therapeutic support through coping strategies or help someone rebuild their life after suicidal ideation or other psychotic episodes. Using referral counts and training data models to recognise people experiencing multiple overlapping

---

[52] Guardianship issues under Section 10 of the Mental Health Act 1983 is currently under review to address complaints that people with cognitive and developmental disabilities are overrepresented amongst those who are involuntarily sectioned and that the legislation needs a massive overhaul to reflect contemporary realities.

vulnerabilities (such as DV and mental health concern) as being at high risk can backfire if there is no concomitant support to offer.

Figure 5: Indicators developed from the Data Warehouse that are fed into the Risk models



In seeking to make mental health issues tractable in this manner, the real risk is that the children and young people who are tagged by this system are treated as liabilities using up too many staff hours. By defining data signatures through these models of ‘people like you’ who are likely to require expensive statutory interventions in the future or liable to take up a lot of staff resources, this datafied system can turn into a self-reinforcing loop of identifying potential mental health cases that need to be investigated – but due to the lack of capacity to provide individualised attention can consequently end up exacerbating their poor mental health and create lasting scars. By trying to pre-empt where case load is going to come from by looking at historical trends and predicting the salient features of risky cases, it is highly likely that the individuals who are targeted with precision would be someone who has had extensive experience with services. Such a person who as part of a domestic violence situation was manipulated into being sectioned or degradingly was put through interventions against their will might find further contact with statutory services triggering if not handled in a trauma informed manner. It would be like calling someone up every day and asking if they are suicidal for assessment purposes, triggering them at regular intervals and subsequently driving them over the edge.

Tools that data scientists have built into the system cannot distinguish between genuine need and data artefacts and cannot incorporate variations in vulnerability. By generating predictive scores out of a managerialist desire to prevent liability or culpability if a child in custody of someone with a mental health need dies (through neglect or abuse), these systems are unable to incorporate protective factors and user-centred social care relationships that are strengths-based and not based on a deficit model that merely seeks to ascertain if someone is a risk or not. At the time of the research there were no benchmarks or formal guidelines that clarified that mental health or other health datasets should not be combined with those held by enforcement and police departments. Neither were there any systematic attempts to conduct impact assessments of the implications of letting data lead and finding the ‘needle’ or the data signature of a vulnerable person at risk of exploitation or abuse; let alone follow up research into the wellbeing outcomes of those who were targeted and offered statutory services based on these datafied models. In building technical datafied models to target attention at specific kinds of potential victims, the practice orientation is on processing, classifying and prioritising cases rather than promoting an in-depth understanding of someone’s needs and actions. This corresponds to what Social Work scholars have been warning for over the last two decades that the logic of care within the profession is worryingly shifting to one centred around risk as a ‘first-order construct’ (Griffiths 2000) against which frontline worker time and resource is rationed. When the practice orientation is around mitigating risk, the language of managerialism rather than compassion tends to dominate.

Secondly, a large component of the training dataset built from the complete data warehouse is the ‘Troubled Families’ dataset held by the local authority. Much ink has been spilled in highlighting the ideological basis of the Troubled Families programme which began as a solution to the 2011 riots in England; built on the assumption of a ‘lack of moral character among the rioters caused by poor parenting[53]’ (Pleace 2007, Cairney 2019, Lambert 2019). It has been noted that the very institution of the category of ‘troubled families’, following the

---

[53] Riots Communities and Victims Panel (2012)

2011 riots represents an escalation of state intervention into families deemed problematic. The Conservative government's intrusive ideas about the ideal of a heteronormative family feed into these policy interventions in which Ministers asserted a major social problem and initiated a massive roll out of targeted interventions to identify and monitor the behaviour of these troubled families[54].

At the centre of this policy response was the assertion that a relatively small number of families are the source of a large proportion of the problems in society and that this can be addressed by intensifying state intervention into the lives of these families where parental worklessness, criminality, child truancy and anti-social behaviour was allegedly being passed down generations. It was a call to find these households and families and intervene early into their lives so as to prevent social breakdown and the blight of riots and looting on the streets. Anecdotal evidence about 'worklessness', 'welfare dependency' and 'feckless parenting' was activated to make the case that these markers of deprivation were leading to undesirable moral codes, lack of self-restraint, and behavioural issues like addiction and alcohol abuse being passed on through the generations. This framing necessitated earlier interventions into the lives of teenagers in these types of families while simultaneously obscuring the structural and socio-economic causes of poverty and disenfranchisement that led to the 2011 riots. The initial target of almost 120,000 families was speculatively worked up by the government based on previous Cabinet Office estimates going back to 2006 that about '2% of families in England experience multiple and complex difficulties' (Kendall et al. 2010: 1, Social Exclusion Task Force 2007: 4, NAO 2013: 5, Hayden and Jenkins 2014: 635). Each local authority in the country was then tasked with identifying 120,000 families based on select criteria; thereby conflating variables used to identify families in need of support such as the mental health of the mother and sanction such as anti-social behaviour or criminality (Cairney 2019: 7). Conservative led governments expanded the cohort of households councils have an obligation to identify and work with from

---

[54] Within one week of the riots, the then Prime Minister David Cameron (2011) linked behaviour directly to 'thugs', immorality and family breakdown - "people showing indifference to right and wrong…people with a twisted moral code…people with a complete absence of self-restraint…We've known for years that a relatively small number of families are the source of a large proportion of the problems in society. Drug addiction. Alcohol abuse. Crime. A culture of disruption and irresponsibility that cascades through generations. We've always known that these families cost an extraordinary amount of money…but now we've come up the actual figures. Last year the state spent an estimated £9 billion on just 120,000 families…that is around £75,000 per family."

120,000 troubled families between 2012 to 2015 to from to 400,000 after those years with a third phase beginning in 2017 (DWP 2017).

The early modelling defined troubled families as those with poor housing, low parental education, unemployment, chronic illness or disability of either parent, the mental health of the mother, an income below 60% of the median, and an inability to buy certain items of food or clothing. The brief sent out to local authorities demarcated inclusion criteria based on whether the household has an adult receiving out of work benefits, or a child with a permanent school exclusion or suspension or 15% unauthorised absences for over three consecutive terms; or families with a child who has committed an offence or is subject to an anti-social behaviour order (ASBO). These are clear signifiers of complex, overlapping socioeconomic disadvantage. In training models to detect correlates of high probability and high consequence risk on the data on welfare history and on digital transactions that troubled families have had with schools, court system and unemployment services, automated risk categories are being built on systems of exclusion already in place thereby exacerbating them.

These algorithmic determinations of ‘at risk’ categories trained on precise ways of targeting ‘people like you’ means that this datafication process reifies and solidifies their history of interactions with welfare agencies and problematic behaviour is associated with certain types of families. The desire to target interventions based on policy objectives has been in place since before the Troubled families programme were set up and consecutive Conservative governments set measures in place to automate austerity. Even during the New Labour years post 1997, the infrastructure was in place to blame the post-war welfare state for increasing welfare dependency. There was a pervasive rhetoric around an ‘underclass’ that was unwilling to work, dependent on welfare benefits by choice and left behind by economic changes that required a different set of skills (Welshman 2013, Carter 2021). This led to a plethora of data processing technologies focused on preventing fraud; which was imagined to be all pervasive with ‘work shy’ benefit cheats conspiring to defraud the welfare state at scale. Enforcing conditionality and an increased surveillance of behaviours of those on benefits was also the order of the day. The Blair government was focussed on getting as many people as possible back to work; and there was a decisive focus on socially excluded people and on incentivising lone mothers back to work[55] by increasing their compliance requirements (Pleace 2007). There

---

[55] This focus on getting single mothers working was not one that was pushed by previous governments.

was an accelerated effort to use technological tools to define socially excluded groups more precisely for preventative interventions; and individuals and households who fall into this group needed to be identified and tracked (Carter 2021). Henman and Adler (2003) posit that the development of technological capabilities to analyse ‘welfare client’ statistics has led to subpopulations being constituted as risky and problematized as objects for governance. When datafied models are introduced into a system that is built to discipline marginalised people back into work in the formal economy, the allocation of risk derived on measurable types of characteristics of individuals are projected onto to groups and subpopulations who then have no means to contest them.

The effort of preventing sexual abuse and criminal exploitation or indeed the detachment of young people from the tethers of the schooling system causing them to be ‘not in employment, education or training (NEET)’ is noble and laudable. However, training signs of exploitation indicators on datasets drawn predominantly on those with extensive welfare histories on record reinforces negative stereotypes and reinscribes the conditions of their disadvantage. Most importantly, predictive risk labelling based on these data traces reifies assumptions about being defined by your past when the circumstances of a previous victim is transcoded into a prototype. The training dataset for the Child Sexual Exploitation (CSE) risk thematic is based on care experienced children who were assisted by Barnardo’s. Barnardo’s as a national agency that aims to support vulnerable young people explicitly believe a child's future should never be defined by their past - and in trying to use data to prototype what an abuse victim looks like based on their historical statutory data footprint, this becomes a system that is rigged to become a self-fulfilling prophesy where a certain type of a child from a certain type of household will go on to become an abuse/exploitation victim.

There are clear class biases baked into these datasets. A landmark report by the National Society for the Prevention of Cruelty to Children (Cawson et al. 2000) lays out findings from a comprehensive national study that explored young people's experience of childhood maltreatment by parents, relatives and other carers. It presents undisputable evidence that there is a socioeconomic link to neglect and extreme physical violence; but no discernible corelation between socioeconomic status and emotional and sexual abuse and moderate physical abuse. That is a crucial finding that often gets ignored. The instances of neglect, abuse and physical maltreatment in middle class families would not show up as readily on existing electronic case management statutory systems. Therefore, for all its focus on troubled families, by trawling through datasets in a ‘data-led’ manner the legal definitions and distinctions between severe, moderate and mild forms of physical and sexual abuse are lost.

During initial fieldwork, a female safeguarding liaison officer within the police force described with a degree of anguish how she had cause to go into the home of a family within a very posh neighbourhood and initiate an assessment discussion with a secondary school student aged 10. The visibly wealthy parents were hostile to her presence in the house. During her ‘signs of safety’ conversation with the child it became obvious that the girl was being raised by a non-English speaking *au pair* and saw her parents only sporadically despite living with them. The child had poor language capacities (as she spent most of her early childhood years talking to the *au pair*) and this was being manifested as a distinct inability to communicate her needs in the school and family environment. Despite the police safeguarding liaison officer’s assessment that the child was living a fairly lonely and neglected childhood with obvious implications for her emotional, social and cognitive development, the officer was intimidated into not writing a severe report or prescribing parenting lessons for the parents. The record that would be generated for aggregation into the risk score through the datafied model in this case would not in any way reflect the severity of the potential harm for this child; or indeed for the larger group of ostensibly affluent yet neglected children as there wouldn’t be data on file on them in large enough numbers for there to be a measurable type that identifies their needs.

The NSPCC study and social work practitioner grey literature shows that the scale of middle class child maltreatment is grossly underestimated. If a family is of working class origins and the teenager is acting up, school and social work professionals tend to think the parents are messing up their parenting. This leads to judgements about the household needing statutory support through early offers of help which ranges from universal services such as parenting classes to having a staff member handhold the mother through the aggressive teenage years via programmes like the ‘families in focus’. However, if it is a middle class family the immediate reaction is to get the child to see a psychologist and if referred to a private psychologist there would be no mental health concern record to input into these data driven systems.

In pre-empting future abuse in a datafied manner and looking for hidden risk in the haystack, the system imports biases and amplifies in a precise way the vulnerabilities of young people on the fault lines of society. Especially for those with care experience, these risk alerts, and interventions by the law and statutory professionals become part of the young person’s record and part of how they are known to police, social services, and to their future carers. Research by the Foundation for Information Policy Research (FIPR) commissioned for the Information Commissioner's Office (ICO) argued that in the past children who were from Black Caribbean or Black African backgrounds, or from poor neighbourhoods or travelling families suffered disproportionate police attention because of the expectation that they would be more likely to offend (Anderson et al. 2006). They would be stopped and searched on suspicion of delinquency more frequently. Concerns about aggregating negative indicators from health, school and other records without any capacity to code in protective, strengths based features were raised twenty years ago.

> '*A perfectly law-abiding youngster from a difficult home background, who has perhaps struggled to overcome learning and health difficulties, may find at every turn that teachers expect less, and that police attention is more likely. As the causes of this discrimination are online, the youngster cannot mitigate them simply by dressing neatly and being polite. The data and algorithms used as a basis for discrimination might not be accessible to the victim… and thus a victim of unjustified discrimination might end up with no recourse.' (Anderson et al. 2006)*

Warnings were sounded about the foundational appropriateness of collecting, processing and retaining data through pervasive children's databases in the public sector from a privacy and data protection perspective before the growth of this level of computational capability and the ability to pre-emptively make inferences about problematic behaviour. When digital data signatures and proxies are made for 'at risk' patterns so much so that a young person cannot mitigate these labels by dressing neatly, being polite and performing well at school it raises serious questions about whether intergenerational deprivation is being criminalised. There is scarce awareness of the ways in which drawing on the architecture of the data warehouse and letting 'data lead' affects the articulation of the outputs of the model and subsequent early markers of exploitation.

This overriding desire to pre-emptively identify harm by letting data lead due to a risk orientation geared towards avoiding costly interventions in the future has unintended consequences. There are blatant blind spots that the data scientists or even organisations trying to think about how to make algorithms work for society have only begun to grasp (Ada Lovelace Institute, AI Now Institute and Open Government Partnership 2021, Ada Lovelace Institute 2021). We have seen how the predictive aspect of the scoring system relies on imperfect datasets and data practices that solidify and reinscribe historical interactions with welfare agencies and safeguarding teams. In seeking to define the measurable type of what looks like a victim, these risky features are reified. Consequently, there is no way to tell the difference between genuine need and a data artefact baked into the overall modelling due to a lead professional’s opinion that was clearly based on stereotypes about someone with care experience or based on unconscious bias towards a particular socio-economic group with visible markers of disadvantage.

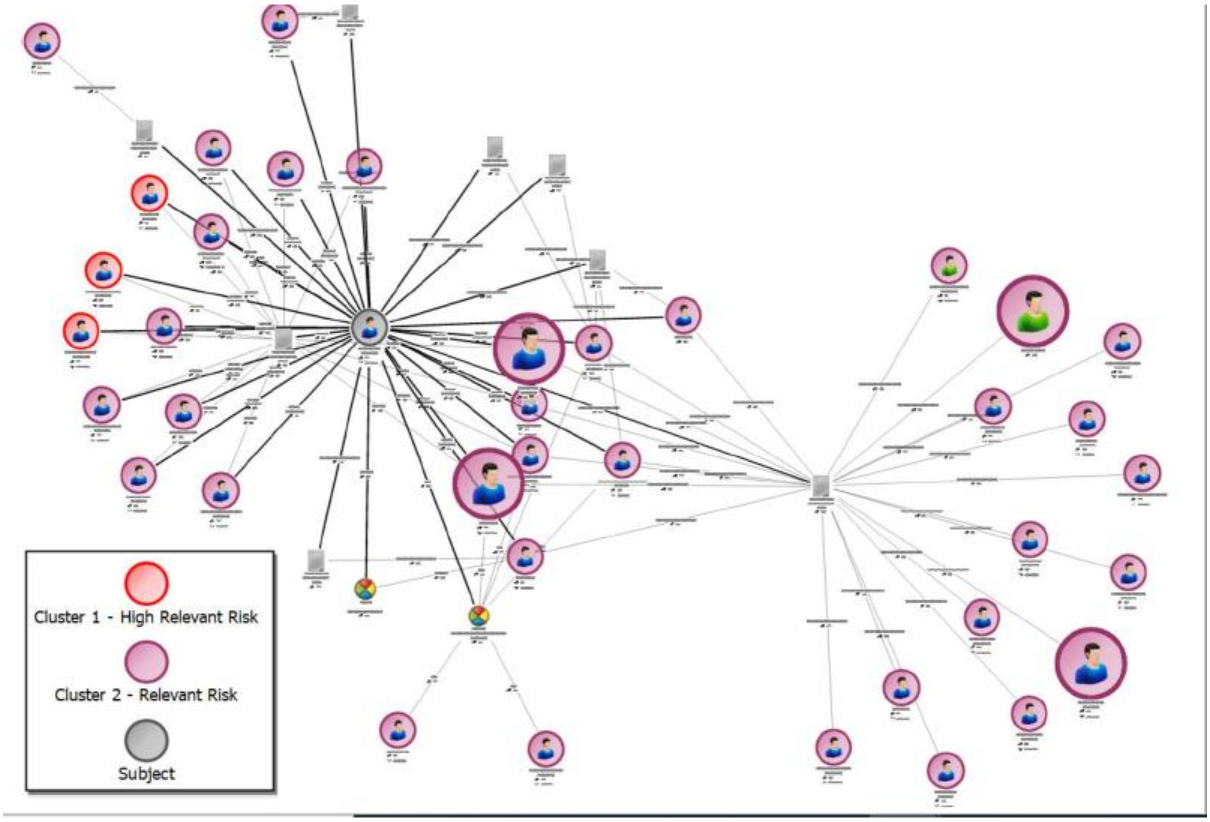
This is the part of the datafied model that the principal data scientist who designed the system feels would be hard to defend publicly as the ethics of algorithmic decision making debate progressively heats up over time. While anticipating a possible future where the public sector decides to not use machine learning tools, he felt the network score as the product of a social network analysis (SNA) output derived through a Power BI data visualisation programme was a good proxy for a child’s social context. The graph database that is the raw input for the SNA visualisation is made up of a very large edges table and the empirical nature of being able to see who is linked to whom in the underlying person link table inspired confidence in the final visualisation output. Being markedly distinct from the machine learning logic of the predictive score, this interviewee trusted the network visualisation to objectively and transparently represent the child or young person’s social context as they still needed to prioritise who they would be working with.

4.2 Peer association risk: Measurable types and attempts at contextual safeguarding

Risk based prioritisation reifies socioeconomic disadvantage into digital proxies and by opting for a data led approach to pre-emptively identify signatures of abuse and exploitation, there are consequences. By defining risk in the manner that Project Vision C does, informational relationships become important and this is best exemplified in the network score.

Drawing on an approach used to investigate organised crime networks, all known linkages around a child on the Police crime and intelligence reporting system (Niche) can be retrieved to explore who knows whom (see Figure 5). Every child gets a score and if they are in contact with someone who has a high risk score, based on an assessment of whether they are friends or associates, the child’s score is bumped up. Using the network visualisation tool, it is possible to zoom in on a child (subject circled in grey), expand all of their associates, and the events that are linked to them as well as the people that are linked to those events. By clicking and going from node to node, it is possible to map almost all known information on a person from the vehicles they are linked to, phones calls, high risk peers, and the places they frequent.

Figure 5: Peer association risk through network mapping



The project proponents are quick to add that for there to be a link, someone has to be recorded three times on the system as either committing a crime together, or being present at the same space or written up together in the same intelligence reports. This part of the modelling corroborates co-occurrences in police systems and concludes that there is a relative link between people or that they know each other. It goes so far as to classify that relationship as a good friend, medium friend, associate and so on.

> *What we were doing was, we have in our data in police systems, you can link Person A to Person B, and you can put a date on that relationship. And you can say... siblings, parents, guardian, gang member, criminal associate, you name it... business associate. There's all sorts of things in our crime and intelligence reporting system. So that gives us a list of who knows who*
>
> *Interview 14, Business Intelligence Developer, Project Vision C, Aug 2021*

Attempts to understand the risk environment a young person is in by forming a full view of the household and transferring risk scores from one sibling to another have previously been met with alarm. This modelling takes this further and attempts to understand the social context of the young person through promiscuous ‘data associations’ (Amoore 2013) that map a potentially endless network of their peers. Informational relationships with others such as phone calls and being caught in an incident together at a logged hotspot is inferred as evidence of a real relationship. If a young person has been stopped and checked with another person, this creates an incident report. If someone has let someone else use their phone and on the police system if they have been linked to the same phone over the same time period, the inference is that they know each other. Durability of these links are important as the logs don’t capture when there is a relationship ending or changing nature – changing phone numbers is often the only time you drop off the system.

Being caught at a park dealing marijuana or being in the vicinity of someone who was using recreational drugs and then being in contact with someone with a high score three times links them; making the flags against someone’s name more dependent on the scores in a self-reinforcing, generative manner. Most of the behaviours flagged by these systems are adaptive coping strategies and by trying to manage them in this manner more harm than good is done. Through a kind of retrofitting, this aspect of the network score was described as ‘contextual safeguarding’ where after building the model they discovered academic literature on how extra-familiar relationships are factored into child protection assessments. Therefore, ties to ‘risky actors’ determines how visible you are to these systems and context is captured into data as informational connections.

Real world complex networks have properties that are not easily captured by a visualisation tool no matter how sophisticated it is. The project proponents see co-occurrence in a police incident report as evidence of a relationship. The person link table generated for the network score is regarded as extremely reliable as the police intelligence analyst viewing the raw data sheet can actually ‘see’ the links. Social network analysis is operationalised differently in different contexts and drawing network boundaries is an incredibly political act of demarcation (Titus 2011, Titus 2013). In the scientific literature, there are long standing and fierce debates around the bridging and brokering functions within a network once it is mapped out. Choice of centrality measures, going even two degrees outside a set of seed nodes produces networks exponentially larger and often in no way related to the seed nodes where the method started (Hogan 2021). How the visualisation software calculates structural equivalence can display someone as being closer to someone than they really are. Drawing inferences based on a link doesn’t necessarily reflect the underlying structure of the child’s peer group as the model supposes. Using a person in someone’s peer network to make a determination about another person that they have come into contact with opens up the social context with endless linkages and an exponentially growing network. When the data scientist in charge of developing this aspect of the datafied model talks about the network score as the most sure component of the composite score, this rationalisation is a sleight of hand that is hard to believe.

This manner of looking at the social world in order to distil them into a few key relationships raises concerns about whether they are really doing contextual safeguarding. Direct and indirect relationships are inferred based on what is displayed on the SNA visualisation tool where records on the crime and intelligence reporting system seems to stand in for the young person or child.

4.3 Domesticating the scoring model

Even as datafied allocations of ‘at risk’ are made reconfiguring the organisation of work in significant ways, this remains only one side of the picture. Social workers and other users of the model in the council have developed improvisational tactics and modifications of the system to make it work for them. These unintended adaptations of the model do not necessarily correspond to the design choices built in and reflect local realities, needs and the sense making activities that surround algorithmic tools. For example, I heard from education inclusion officers and frontline users of the score that the most useful part of it is in how the social worker attached to the child in question is easier to locate. Instead of seeing the risk score as directly measuring the severity of the situation that a young person is in, one of the key uses of this complex record linking exercise was that when there is a call about someone potentially being at risk of harm, it is easier to check if there is a named social worker already working with that person. This made follow up enquiries easier and often this was more valuable than whether there is a score of 40 or a 90 against a child’s name. This works as an indexing system or a shorthand to bridge organisational distances. This final section pays close attention to the manner in which users interpretatively engage with the output of the automated data model and embed it into institutional routines.

While the score erases everything that cannot be quantified by the risk-measurement model, practitioners retrieve these. The voice of the child is noticeable by its absence in these systems. Traditionally, a social worker would sit with a child and through a therapeutic sort of narrative relationship, ask questions about their feelings. The below frontline user admits how hard it is to incorporate the voice of the child in terms of their representation and involvement into a system build around a risk score.

> *I mean, it is really difficult to quantify something in data that's sort of more of a feeling or thought or you know, how somebody is actually experiencing their lives. I think that's going to be... it's always a challenge. I think.*
>
> *Interview 16, Project Vision C, Aug 2021*

In the work of her team, she has retained signs of safety assessments so that there is a way to write in the concerns of the young person that sits behind manifestations of high risk behaviour. Carrying a knife to school would be scored as a high risk. This interviewee described having to find out more about how that young person is feeling, and why they thought that would be helpful for them to carry a knife and writing in the safety concerns of the young person on their journey to school into the assessment. Rather than relying on the score, she opened up the conversation about which neighbourhoods had crime hotspots and asked for further community risk assessments to be done.

Several interviewees admitted that the scoring might be wrong and that they still needed to ‘dig behind the data’. In case conferences, practitioners would manually add names to the risk cohort and talk about those the scoring process had not listed. They would refer to the risk model as a guide and speak about the young people who they needed to speak about. The way frontline users interact with the modelling has changed over time. Someone who was very involved in the early stages of the modelling but had now moved on to a different role in the same council and started going to meetings after a long period of time exclaimed that it was a lot freer and the ‘conversation would flow how it needs to’. Over time, the initial data science team received requests for explainer labels to each score that set out the origins of the risk in sentences. These scores remain sites of tension where frontline social workers and programme managers interpretatively engage with outputs from automated systems and figure out how to push back based on aspects of their professional expertise they felt were being eroded.

In meetings with schools when asking to set up additional support for vulnerable pupils as flagged by the model, council staff were asked what tier they were in. Safeguarding and management staff at schools did not have the means to look through the database. ‘Tier 1’ therefore became an arbitrary signifier of whether they should be concerned about the young person classified as such “by this multiagency department who have the complete understanding of everything that is going on for them in their lives” (Interview 16). There was a lot of back and forth with the school and council staff have tried to use the score to try to open up a conversation.

> *…if we go back to this sort of high scoring, it does offer us an opportunity to widen that conversation with schools to say... Have you tried this, this and this, you know, is there something missing, or something that's been missed from the education perspective? Have you gone to school nurse services? Or have you gone through your Senco? Or have you, in some cases… have you actually asked them to do a reading test on them. Just to say, you know, have you checked that their literacy levels are of a level that you're not pushing them into things... and they are acting out? Because they don't want to be where they are.*
>
> *Interview 17, Project Vision C, Aug 2021*

Especially in instances where schools were trying to exclude someone who was bringing down the overall academic attainment average, council staff use the risk scores to make the argument that compared to other schools in the region, that particular person's score wasn't too high. They used this new space opened up by mandatory meetings to talk about value systems and how getting high GCSE marks were not a priority for everyone. Some families might just want their children to learn to fill in a form or access a job. What was being manifested as problematic behaviour where the child was 'acting out' (Interview 17) and displaying signs of concern, could be due to them having poor literacy skills or having special needs that were not being catered to in the classroom environment. The above quotation of the council staff member using the score to not flag concern but redirect the school to look at whether the child has special educational needs or disabilities that the local Senco (Special Educational Needs Coordinator) could make provisions for is a good example of unintended adaptations of the scoring system.

Rather than data replacing relationships and individualised attention, in this instance of negotiating with the score, the council staff member retrieves contextual information about the child that had been erased based on her experiential judgement and initiates a broader conversation about the child's needs at the school. Using the system allows for these broader conversations to happen and is a good example of what Crooks calls the 'interpretative resistance of data systems' (Crooks 2019). Despite the potential for data driven systems to become inescapable, ubiquitous and encroach into the most private realms of life, recent studies of data systems in urban schools (Crooks 2019) have shown that there is often disagreements over what data could mean and what they could stand in for. In other words, there is often scope for the human handling the data to contest its interpretation, make it work for them and in the process reduce the reach and power of the data system. These acts of negotiation and contests over the interpretation of the scores that have unintentionally developed over time is a form of ‘interpretative resistance’ and it would be incomplete to study the advance of datafication without paying attention to how it is being challenged by the micro level practices developed by child protection staff implementing these systems.

4.4 Conclusion

Drawing on the findings from the local government case study presented through this chapter and the previous one (Chapter 3), we see the messy manner in which the data driven risk scoring model is being implemented. I have demonstrated how social categories such as peer groups are translated into abstracted data categories and the consequences of relying on measurable types of ‘people like you’. The manner in which ‘at risk’ is conceptualised and operationalised in this model as a numeric value built on an index score, a predictive model, and a network measure exemplifies the logic of managerialist objectivism and risk orientation that has been transforming this sector over the last twenty years; now with a data obsessed prioritisation model that values informational connections and the primacy of the administrative record over that of the voice of the child.

Chapter 5

Datafied Welfare Services: Role of proxies and insidious algorithmic grouping

We have seen how ‘at risk’ categories and scoring systems are built to identify and target service delivery at the local government level. Taking a step back from the statutory world of child protection and youth violence reduction programmes, in this chapter and the next, I draw on empirical material from a large digital transformation unit of a central government department with the remit over welfare, pensions and child maintenance policy. Through a comparative case study approach it becomes possible to see the unfolding of datafication in institutional contexts that are similar in that they are both involved in the administration of social care and welfare provision; but reveal different aspects of the growth of data driven technologies.

This department is at the cusp of a once in a generation digital transformation which will touch the lives of over 20 million people in the UK every day. Positioned at the core of every digital service they design is a mandate to change the utilisation of data in the sector through new organisational roles focussed around data science, data engineering, data management and data analysis aimed at ‘making services better for users’. Data driven processes are at various degrees of penetration into the existing departmental culture. There are teams that merely digitise paper forms that have been posted in using unsophisticated OCR technologies, while departmental goals to increase reliance on datafied processes remains aspirational for many benefit lines. However, there are stated ambitions to increase the use of ‘customer insights’ by developing analytical models and embedding them into operational systems where traditionally a frontline staff member would interact in person with a pensions or benefits recipient. Various work programmes aimed at using data to revolutionise decision making in government are underway. These include inserting various risk alerts into frontline digital services and using Natural Language Processing to listen to civil servants placed at different levels at the department in addition to working age claimants and pensioners.

In this chapter I will explore three different instances of datafication – firstly, back end analytics that trace user journeys through the website, then A/B testing and the more futuristic design of the digital service to nudge a recipient based on the anticipation of their personalised needs around life events. Each of these data practices categorise in different ways and reviewing them one after the other helps us see the implications of these categories on vulnerable welfare recipients. I then conclude by drawing on political imperatives as an explanation for how data systems were designed the way that they are in this department. This chapter introduces and develops the role of digital proxies in sustaining datafied administrative infrastructures. In doing so, I demonstrate the institutional embeddedness of processes of datafication; and how organisational culture and mundane bureaucratic rationalities dictate the choices that data workers have to grapple with. In this way, I contribute to the emerging canon of Critical Data Studies that has called for a deeper understanding of the sociotechnical issues that are entangled with datafication by locating practices within the social, political and economic conditions of the creation and use of data (Dourish 2016; Seaver 2017).

Data from dwell time on a specific site on the departments' webpage, user movements through forms and other inferences from google analytics is used to personalise the service by forming a view of the welfare recipient. While these data driven technologies promise to not base decisions on characteristics protected under the Equalities Act or subvert the privacy mandates of data protection frameworks, they group and divide people within social service delivery in remarkably new ways. Drawing on Wachter's (2022) work attempting to close the gap between legal doctrine and emergent forms of algorithmic discrimination, the final section of this chapter raises questions about whether these taxonomic systems even have the capacity to trace the unintentional redlining perpetuated by their datafied processes. Categories and groups that are generated through data driven automated classification and user segmentation do not currently enjoy legal protection unless they can be mapped onto an existing protected group and such linkage is rare in practice. In addition to matching data points on a citizen held by different governmental agencies, there is a stated aim to use data from the service in order to learn about the user and develop proxies for suspicious behaviour. Inferences drawn through these proxies though well-meaning cause allocative harm that is hard to even map out.

After introducing their overall digital vision for hyper-personalisation and a 'join-up' around life-events, in the next chapter, I explore how an overriding heteronormative ethos centres some people's experiences and optimises the service for them. Drawing on interviews and documentary sources, these two chapters raise questions about digital exclusion, the segmentation of welfare recipients based on their behaviour on the website and the broader implications for when the traits of those excluded by data-based categorisation cannot be mapped onto a protected group. Instead of a single datafied model that has been embedded into the workflow like in the last case study, here a series of analytics processes instituted by different teams across the department are considered. In this regard, I adopt a lateral rather than an in-depth analytical approach for this case study.

5.1 User Journeys: Backend Web analytics and Data Trace Capturing

At this department, there is an ongoing push to embed analytical models into operational systems with data being continuously captured from the service to generate granular, real time insights about patterns and risks. By tracing browsing movements through the departments' gov.uk interface, it is possible to form a view of the welfare recipient and draw various sorts of inferences.

Metrics on dwell time on a specific webpage is captured as is data from someone's movement across webpages while they complete a form online. User movements through forms and other inferences from Google Analytics is used to develop measures of 'customer closeness' (Interviewee 3) and computational guesses about user or customer intention. This means that certain kinds of 'back and forth' behaviour between webpages or keywords or metadata that looks like fraudulent behaviour makes one more visible to these systems. There are a range of insights that can be drawn from the user interface as most digital forms and applications are now online on the gov.uk website. These include completion rates, page view counts, referral where a claimant accesses a website using a link from another website and bounce rates which are the percentage of users who leave a site after viewing a single page without triggering an event.

> *…if people are taking a long time on a set of questions, then maybe the questions are being asked in a bad way, maybe they're asking for information that we don't necessarily need. Alternatively, there is a... if people are getting to two pages into the form and then giving up then that maybe there is a problem with the type of data we're asking them for at the start of the journey. So if we're asking people to enter information about bank accounts, and for health insurance policies, and all that kind of stuff, if we ask them to do that straight away, then some people will just think I can't be bothered with this, you know. Whereas if we ask people to enter information in a way that's comfortable for them, then maybe… we get ...success.*
>
> *Interviewee 5, Software Engineer, November 2020*

A senior data scientist with a conversion rate optimisation[56] background spoke about how dropout rates midway through an application are captured as they particularly want to reduce the number of people who give up on the digital form and make phone calls to the call centre seeking help**.** In addition to the above interviewee's concern about someone taking too long on a set of questions, inferences were made about the kind of user for whom the section of the application was hard to get through. Excessive 'backwards and forwards' movement through different screens in an application was seen as a proxy for the wording of questions being distressing for the applicant. It could just be that the claimant was clicking through to see what documentation is required on the next page. Regardless of the subjective singularities of individuals, their unique circumstances, motivations or intentions (Cheney-Lippold 2017: 5), their browsing movements are being interpreted and the user is being computationally calculated and classified a certain way.

Proxies for user behaviour are being built based on which part of the application process takes longer or slower for different people. Based on data that is collected at a point of the claim process, computational guesses were being made about whether people with a particular health condition take longer at a certain part of the process so that they could be routed differently in the future. Disaggregated data based on location and browser type is used to understand if 'somebody in a different part of the country suffers a different way' and 'in what ways are we or our services either advantaging or disadvantaging people in their journey through the digital service?' (Interviewee 11). There is an ease with which the context of someone's health condition as slowing them down is interpreted as data movements that are neatly encoded, defined, and wrapped up when it shows up on the analytics dashboard.

With the introduction of GDPR in 2018, this ability to track users was hindered causing mild panic among these data analysts about consent models that were at the heart of their analytics system. I was told that around 80% of claimants had clicked 'agree' to the collection of cookies during their use of the website. These inferences from dwell time, what 'looks like' someone in distress or proxies for fraudulent behaviour are fed back into the service in order to make it more nimble, agile and responsive. 65% of gov.uk use was from mobile phones and for people

---

[56] Conversion rate optimization (CRO) is a web analytics practice from online marketing of increasing the percentage of users who perform a desired action such as purchasing a product, clicking 'add to cart', signing up for a service, or clicking on a link.

on low incomes, a majority of them used an outdated android phone rather than the latest iPhone. It could just be that the website froze due to the phone quality or a quirk but this clicking behaviour is thought of as being able to tell them something concrete about the user. A datafied view of the claimant is being built from the user journeys and there is a stated aim to increase the capture and use of data from the service to learn about the user. User researchers (who are dressed down social scientists) attached to the team might then investigate if the wording or formatting of the page was slowing people down. I was told that they actively look around on internet support groups and blogs frequented by welfare claimants and the information provided by NGOs like Citizen's Advice to get a sense of common problems.

These proxies built on data signatures are problematic as certain kinds of online behaviour is being coded as risky behaviour that requires further investigation. Similarly, excessive dwell time on a particular page is being read as indicating distress and needing referral for extra support. As will become evident over the course of this chapter and the next, these techniques do not take in account digital literacy issues, or the full, long term implications of segmenting vulnerable low-income people who have to use the compulsory website in order to get a service.

5.2 A/B Testing, Large scale data and User Experimentation

> *So, I guess the purpose of our team, one of the reasons why we use A/B testing is because we provide answers and evaluations to guide the development of the service. So, it may be the case that in any moment in time, say, we have to implement a change to a [working age benefit] because of a policy that has been agreed by the Minister. There may well be multiple ways in which to deliver that objective. There might be pros and cons to doing it in different ways. And one thing that data science offers is the ability to compare those options analytically and help shape the evidence as to which one to get. So, you know, we might have to communicate to a claimant some change in the way that it works, you know, some, some change in the way that their entitlement changes. (Pause)*
>
> *Interviewee 1, Senior Data Scientist, Sept 2020*

A/B testing is another significant data practice that is used by this department as a probe to understand user behaviour in real time as well as to compare options when settling on the best way to implement changes to the digital service. A/B testing is a technique that randomly splits benefit claimants into groups that receive different variations of a feature of the digital service. This technique is possible in the first place because this Department's major welfare-to-work benefit is now delivered in a predominantly digital environment. This means that by capturing uniquely indexical data on claimants based on how they interact with job coaches and access support, it is possible to compare various pathways through the service. Through this randomised experimentation process, two or more versions of text on a website explaining a feature of the claims process are shown to different segments of visitors to the website at the same time; to determine which version is easiest to understand (or achieves a particular predefined outcome). It is possible to demonstrate if variations of the positioning of the text, font, or the navigation flow through a series of webpages or other such changes to the content of a page can perhaps save five minutes off a phone call to a staff member at the job centre or a call centre.

Going beyond readability of the text, style, content management and the website optimisation process, these trials are frequently run in parallel and at a large scale to measure variations in real time to features of the service that are of greater consequence. This includes iteratively testing different ways to verify a claimant's identity and monitoring the data as it accrues. For example, the same data scientist described how results from A/B tests were used to check if two factor authentication would slow down or increase the dropout rates of those registering and logging into a Universal Credit account. Two factor authentication, introduced as a new process to replace the Personal Security Number, meant that claimants would need to register a mobile phone that would be used to verify who they are at key stages in the Universal Credit payment system. At the point at which this new functionality was added to the claims process, data scientists sought to accumulate evidence to learn whether it prevented people from creating their Universal Credit account and whether it resulted in additional operational cost due to extra contact from those struggling to register their mobile phone and failing to sign in[57].

---

[57] These details that complement my interview and participant observation information was retrieved from the response provided to a Freedom of Information Request (documents 76641_CYI and 76641_2FA) submitted by the Information Rights activist Phil Booth following a presentation at the Institute of Government that is archived on YouTube. It was submitted through the ***whatdotheyknow*** website. Additional operational cost due to extra contact from those struggling to register their mobile phone and failing to sign in meant the extra time spent by call centre operatives who would have to handhold claimants facing difficulties. The full list of trials, captured metrics and detailed results show how they conclude that two factor authentication did not lead to significant drop outs. The document shows how they capture small signals in data but miss the point as there is no means to know this for sure or check if the drop outs were due to claimants not having an email address or a device. Access to a device, email and digital literacy is important as typical claimants are from a low socioeconomic background. The percentage changes in users that log in are captured without enough disaggregated data that lets one investigate why there are a change in users; making the conclusion highly suspect.

In this instance, the A/B test was designed by continuously allocating 11,000 claimants into a 'treatment group' with the new two-factor-authentication feature enabled on their accounts. Concurrently, another 11,000 claimants who were attempting to complete their registration through the existing claim process were tagged as the 'control group' and their progress was unobtrusively tracked, traced, and measured.

Before going into the technical details of the trial implementation process, the interviewee prefaced his explanations by saying that the welfare platform he had responsibility for began as an agile product that was released into the live environment as a minimum viable product with much functionality still to be added. Another interviewee working within the analytics team of a different benefit line within Department A described an 'A/B/n' test where three partially developed versions of a child maintenance payment application form was shown to one third of their users simultaneously in order to gauge which design[58] to invest in and develop fully.

This 'test and learn' approach requires iterative tuning and optimisation of the digital service based on large scale data from user experimentation. There are clear advantages to testing prototypes on real users to gather immediate feedback - it can open up possibilities for reassessment and re-design. Results from large randomised experiments become available almost instantaneously and allow system designers to make causal conclusions about changes made to the digital service and user reactions to them. These sorts of quasi-experimental methods which include classic A/B testing designs have been normalised in the technology sector with many regarding them as necessary to improve products and services (Kohavi et al. 2013, Bird et al. 2016, Fabijan et al. 2018). Computer scientists releasing these experiments into the wild see themselves as working on systems, digital architecture, and mathematical optimisation models rather than on human subjects (Metcalf and Crawford 2016). In this case however, a large number of people depend on this critical digital infrastructure and the 'user' being experimented on is a welfare recipient navigating this system at a vulnerable point in their lives when they have just lost a job and need to apply for income support. Serious questions arise about the epistemic validity of this sort of evidence that is experimental, based

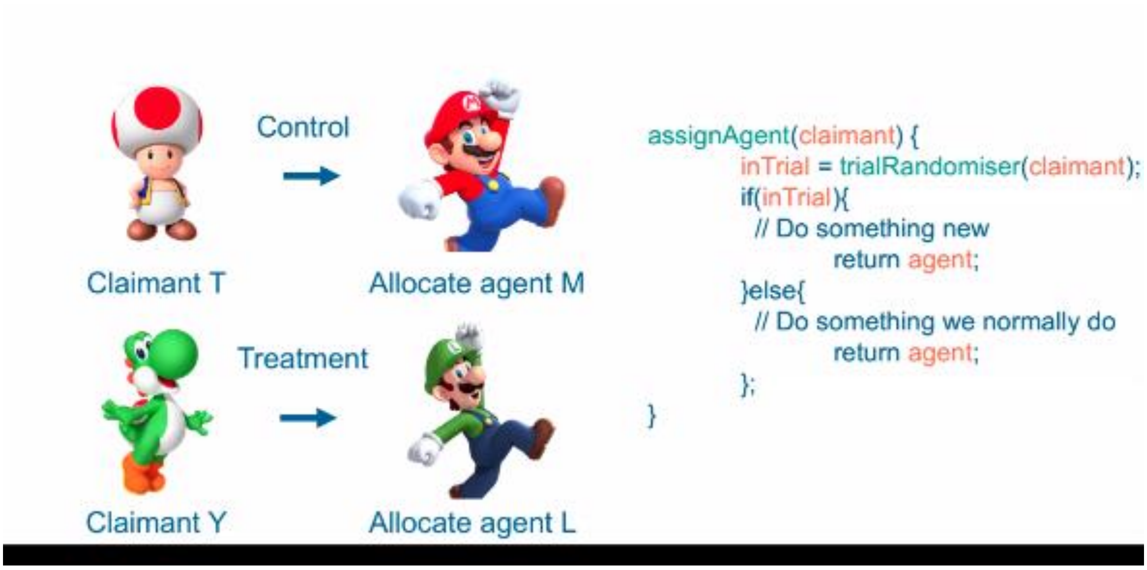
---

[58] This was aimed at working out whether the different circumstances of parents who share custody of children were captured in the wording and text input options in the form – this kind of testing can be seen as a way to gather 'user need' in real time.

on live users logged on to their accounts at any given time and created through the real time validation of design choices where vulnerable users are unaware if they have been enrolled into an experiment. There is also no user agreement that specifies that these sorts of large-scale testing would be conducted at any stage of one’s engagement with the department.

A/B tests were developed in the marketing and e-commerce sector within ad-placement systems that are geared towards encouraging click through rates and nudging a user to make a purchase. The troubled genealogy of these trials and the contradictions that arise from marrying these methods to a sensitive policy domain is most evident in the illustrative example and heuristic used by the interviewee below (Figure 6). The below allocation logic forms the centrepiece of the team’s automated A/B testing framework in the ‘real business context’ of administering a major benefit. Figure 6 hypothetically details how a claimant would be assigned an ‘agent’ who in this case is a frontline member of staff or a job coach through a trial randomiser method. Using the java function designed in-house (displayed on the right in Figure 6), a claimant is assigned ‘Agent M’ in the control group and the next claimant ‘Agent L’ in the treatment group based on whether the claimant is in the trial or not. This is a simple piece of logic that can be expanded upon, and complicated tests can be run using this branching functionality.

Figure 6: Automated A/B test to hypothetically route a claimant to different frontline staff in the control group and treatment group



```
assignAgent(claimant) {
        inTrial = trialRandomiser(claimant);
        if(inTrial){
          // Do something new
                return agent;
        }else{
          // Do something we normally do
                return agent;
        };
}
```

Source: From slide deck of presentation attended during participant observation, Sept 2020

By establishing a formal architecture for facilitating random allocation to a different frontline agent or staff member, it becomes possible to distinguish between claimants, tag them and route them through different control and stimuli variations to measure differences in outcome (Figure 6). This means that some benefit claimants will experience inconveniences and risks that others don't. The above interviewee is attached to a modular team that exists solely to measure such differences in a very precise way and then to compare, evaluate and provide advice on what he believed was an "immediate, data driven way to proceed".

Discussions of A/B testing within Critical Data Studies are centred around the ethics of experimentation with one controversial case featuring prominently. Facebook's emotional contagion study in 2014 caused public furore when it used A/B tests to see if the positive or negative tone of one's news feed would trigger similar emotions on a large scale (Crawford 2014, Kramer et al. 2014, Grimmelmann 2015, Meyer 2015, Meyer and Chabris 2015). While the uses of these methods in closed digital bureaucracies aren't as well known, the key issues revolve around the ethics of knowingly distributing different outcomes to those in the control and treatment groups and whether we can be certain that the conduct of the trial itself will not expose participants to unfair risks. In the literature, this dilemma of experimenting power in platforms is exemplified by the figure of the 'explorer' in navigation systems (Bird et al. 2016, Rosenblat and Stark 2016). Apps like Waze and Google Maps provide real time directions to millions of users by providing suggestions on routes that save time based on the current traffic conditions. In order to gather information on routes that were previously slow, these platforms deliberately send some users along it to explore if conditions have changed[59]. Bird et al. (2016) describe the cost of providing some users (and not others) with suboptimal outcomes in a manner that leverages their ignorance and potentially violates their expectations and preferences:

> *'For some users, taking a slow route might mean that they are slightly late for work; for others, though, it might delay a trip to the hospital… A person who will lose their job if they are late to work might decline an invitation to serve as an explorer, even if there were some chance that they would arrive at work earlier than expected. Other users might prefer to take a route with an uncertain duration over one that is certain to take a long time. Respect for persons dictates the importance of allowing users to make such decisions for themselves in an informed manner - that exploration will discover*

---

[59] The logic of this kind of user experimentation draws from the learning systems that are used in Google. Most search words on Google the database have never seen before so they are constantly trying to understand new queries and build to the body of keywords. This represents a way of building the database through user generated knowledge in real time.

> *information that improves the system as a whole may not justify the risks for any specific user'. (2016:3)*

It is worth noting that for the duration of their engagement with Department A, claimant Y and others who are being exposed to a treatment effect won't know if they are in a trial and have no means to provide informed consent or opt out. The sensitivities in implementing such an A/B testing logic are obvious in a portfolio that provides support in situations as varied as children in low income and separated families, working age citizens through ill-health, disability and unemployment benefits, retired citizens who have reached State Pension age and families who have had a bereavement.

5.3 'Benevolent datafication': Life Event centric service redesign

While data practices like A/B testing and backend analytics to develop a view of the claimant have currently been perfected by various in-house teams, it is instructive to go over the futuristic vision for the department as articulated by high level digital managers. Interviews with senior management show how they aspire to use data to provide a holistic picture of the citizen by joining up various benefit lines and pre-emptively nudging them through various personalised options based on life events. This is seen as simplifying the user's interaction with government and in this section, I present these themes from high level managers setting the goals for the department and visualising how things could be better. Most of these remain hypothetical and aspirational at this stage although in recently built benefit lines that were designed as digital from scratch, it is indeed possible to make these sorts of inferences.

A key contribution of this chapter is the following analysis anticipating some of these problems if this vision for a fully linked up, pre-populated automated service is implemented without any changes. Personalised nudges based on data as a person navigates through the website and their services can raise serious concerns and questions about how people are being categorised and how those categories impact them as they navigate the system. I am anticipating these future problems with this hyper-personalisation agenda which is likely to grow in salience as there are currently no enforceable regulatory guidelines on secondary uses of data in terms of what not to link, how long to keep data points and the limits of making inferences from aggregate information.

A overarching vision articulated by senior members of the digitalisation team is one of using data to personalise the service provision to consciously improve citizen experience and automate eligibility notifications.

> *We went through quite a bit of that. Move from treating the citizen as a case – you are a case to be processed and paid – to treating you as a person, a person who has a birth, death and has employment and has needs. And we are trying to understand you as a real person…* ***almost make decisions before you engage with us [using data], you might be eligible for carer's, you might be eligible for tenant's allowance, you might be eligible for universal credit…*** *and even more. So here are the barriers to access for jobseeker's allowance or universal credit and use data to support that… a dataset will help us say that you should apply for UC because of these data points...*
>
> *A lot of fraud is accidental. HMRC can use the data [it holds on me] and not rely on me to tell them that I have crossed a threshold and am now not eligible for child benefits.*
>
> *Interviewee 2, Senior Delivery Manager, October 2020*

The welfare benefits system has undergone significant changes since its golden period of expansion in the years following the 1945 post war consensus where there was a broad agreement among all political parties that Beveridge's welfare state recommendations and a mixed economy would keep inequality in check. We have since seen the systematic retrenchment of the welfare state, understood in terms of a commitment to universal access to health and education, decommodification which meant social entitlements and citizens' degree of immunisation from market dependency, and the dismantling of social solidarity. In the last 78 years, ideas about those on benefits needing to demonstrate their deservingness for taxpayer funded support began to take root - as did an all-encompassing obsession with preventing fraud where 'work shy' benefit cheats were allegedly conspiring to defraud the state if left unchecked.

Therefore, the above vision of personalisation delivered through bolstering the capacities of the welfare system through the benevolent use of data to form a comforting view of the real person behind the case marks a clear softening of the normative push to stricter regimes of conditionality that strip benefit claimants of human dignity. The manager above is almost toying with the language of universalism[60] where everyone has access to welfare through the

---

[60] The prevailing logic of the welfare state is now a far cry from the initial animating vision of guaranteeing universal access – a large number of tasks performed by staff involve determining and verifying eligibility. This is because in contrast to universalism, selective access to welfare means

recognition that everyone has birth, death, employment and needs; and during their life course everyone interacts with this department and other government departments leaving data trails. Instead of drawing on the guiding principles from the older golden period of the welfare benefit system, he seeks to initiate a shift from a transactional focus by invoking the ability of data to generate a holistic picture of the citizen by anticipating personalised needs and reconfiguring their interactions with government for support. In creating systems that generate pre-emptive prompts based on data points that let the citizen know that they might be eligible for carer's allowance or that they should apply for universal credit, he goes so far as to subtly reimagine the welfare functions of the department through data.

The ambition is that bureaucrats processing cases would look at the dataset from the digital service for verification rather than asking the claimant to send in multiple versions of paper evidence. The target operating model in this case is a digital one with the aim that when they design services now the first thing they should do is design it as though no agent or no person would be involved.

This is an automation as default mindset which assumes that a benefit line can be automated and processed as a 'straight through experience' and then explores what would happen to the workflow if people are added in and automation doesn't work. Designers, in order to understand the 'art of the possible', start with the automation strategy first and then think about which part a welfare recipient might struggle with. They then consider what mitigation strategies to put in place if they know that the digital propensity is low in a particular area. Across interviews, a success story which was reiterated was that of getting state pensions delivered as a straight through digital payment where the front end interface linked straight to the back end, did a calculation of eligibility and made a payment without any person being involved. The ostensible aim is to automate simple tasks freeing up agents and frontline staff for tasks requiring empathy and judgment which are referred to as dealing with the 'complex vulnerabilities' of those who are 'furthest from the labour market'.

This whole enterprise is aimed at providing a seamless, 'joined up' experience to enable a kind of humane anticipatory governance where the system would check someone's national insurance number (NINO), bank details and records held by government on birth, death and employment to generate certain prompts. The interviewee quoted above sees this as a systemic

---

needing to demonstrate through increasingly stringent means tests and the like that the claimant qualifies for support.

change from the transactional focus of delivering direct payments to someone as a case to seeing them as a real person with multifarious needs. Personalisation as an organising logic that gained currency in the 2000s was meant to disrupt traditional modes of social service delivery which allegedly produced dependency of individuals rather than promoting independence where citizens were empowered 'to shape their own lives and the services they receive[61]' based on their needs and preferences (HM Government Policy Review 2007: 7). Reshaping service provision to move away from a 'one-size-fits-all' approach was meant to centre the individual service users rather than the services themselves as the public sector had previously done (Leadbeater 2004, Needham 2011).

In seeking to deliver this personalisation agenda through data points in a complex policy portfolio that touches on sensitive social care and welfare-to-work issues, there are a range of assumptions and presuppositions made in the process of 'understanding you as a real person'. In almost making decisions before the welfare recipient engages with the department through pre-emptive nudges towards certain options, data scientists are building the datafied infrastructure in a certain way. The intentionality, direction and scale of this change is obfuscated when use cases like state pension provision where 80 per cent of claims are now[62] made online is cited as a success story. The key performance indicators of this digital group are percentages of digital uptake. There is an overarching drive to push people online by changing the service availability and procedure – therefore assuming that automation by default necessarily improves citizen experience is a fallacy as they will obviously have more digital

---

[61] This included a lot of rhetoric about personalised services giving choice, control and independent living back to the user. For example, disabled people who were eligible for adult social care were able to access a cash sum instead of directly provided services. They could then design their own support by hiring personal assistants or sourcing mobility solutions from the private sector or a voluntary agency. However, observers (Scourfield 2005) have noted that this shift is merely a symbolic one and a 'qualified form of empowerment' of disabled voices as they do not alter the basic needs-based and means tested basis of the English welfare system (Dickinson and Glasby 2010, Ferguson 2007, Glasby and Littlechild 2009). Onerous and scrupulous financial monitoring of funds dispensed in this manner was coupled with discourses about the personal responsibility of citizens to care and provide for themselves in an active manner rather than be passive recipients. Scholars like Ferguson (2007) argue that the activation of personalisation logic leads to the de-professionalisation of social work. In imagining an empowered adult social care user able to employ personal staff to help with daily living or with getting around, this agenda neglects structural issues of issues of poverty, inequality and the quality of service provision deteriorated drastically now that private companies were involved.

[62] Figures from August 2021 interview

uptake if they direct people to it. Nonetheless, this is seen as a best practice example of ‘using data from other departments more intelligently using automation where it is safe to do so’.

In routing the claimant to zero touch, self-service options, a massive change in the process is unlocked. Those closest to these automation processes seem to be unaware of the scale of this change as they are in the middle of it and the effects of their work are intangible to them. The state pensions team wrangled data fragments about an individual pensioner from datasets that were previously held separately by querying the HMRC. Processing pension requests were seen as relatively simple because for someone who hasn’t ever lived abroad, it is just a matter of extracting National Insurance Contributions, sending digital invites based on age categories and verifying online applications with a mobile or an email code. This seemingly simple automation task meant the ‘customer’ would get a personalised ‘slick experience’[63] when applying and they would not have to wait too long to talk to a call centre agent if they needed to. Automation is at the heart of what this team was trying to achieve and removing the complexity out of previous pension application processes by using data from the HMRC was the epitome of this vision of benevolent datafication. Managers with a planning aspect to their role boasted of how in automating simple procedures, frontline operations staff were spared having to do a repetitive task over and over again while still being able to speak to people who they can really help and thus make a difference. This supposedly increased their job satisfaction and morale.

However, what seems logistically simple in getting older people to go digital isn’t as straight forward as 26% of people aged 75 and over don't have any kind of home internet connection. Latest findings from Ofcom’s Communications Market and Digital Exclusion report (OFCOM 2022, OFCOM 2022) show that for the age cohort accessing state pensions for the first time (ages 65 to 74) at least 8% of households don’t have a home internet connection which equates to 2.27 million households[64] across the country. Among all working-age (18-64) people in DE households (semi-skilled and unskilled manual occupations and the unemployed) the incidence of not having internet access at home was 5% which increased to 30% among those aged 65+. Class origins and disadvantage exacerbates digital exclusion which governments and other

---

[63] Informal conversation following participant observation of a public facing event showcasing cutting edge digital skills in the department. July 2021

[64] Plus or minus 200,000 – OFCOM estimation based on survey sample composition

organisations have been trying to mitigate since the beginning of the 'digital government' project almost 25 years ago.

Systematic inequalities remain in both access and use of the internet. Findings on who uses digital government from the OxIS 2019 survey, which remains the most comprehensive longitudinal research into the demographics of internet use with a nationally representative sample, show how e-government remains the province of people who have higher incomes. When compared to respondents with incomes less than £12,500 per year, respondents in the income categories above £20,000 per year are more likely to use the internet to access central government services, central government taxes, local government services, local government taxes, various forms of policy information and school information (Blank 2020). There is a consistently negative correlation between age and use of e-government for information and services. This speaks to the continuing, overwhelming discomfort of the elderly in managing their pensions online despite being a group that has been forced (or at best nudged) to get their social security provisions online.

Therefore, there is an implicit conceit in plugging into a discourse of personalisation when combining taxation data from another department with a separate public service mandate, work ethos and wrangling disparate datasets to process pension requests as a 'straight through experience'. The foundational rules of granting someone a pension is being changed even as this process is presented as an effort to reduce complexity. In addition to digital access issues as seen in the disaggregated internet penetration figures, there are serious financial exclusion problems for older people that are worsened when automation and data matching is used. Telephony and offline services remain for those who need to apply in a non-digital manner. However, it is disingenuous to present the foundational procedural changes activated by data wrangling processes that extract data fragments held on a pensioner from disparate datasets, as improving citizen experience and providing empowering, personalised direct payments that restore choice, control and independent living back to the 'user'.

Charities like Age UK have repeatedly presented evidence to parliamentary consultations on how there is a form of 'creeping financial exclusion' for older people where financial services such as banking are not equitably distributed or accessible for the 65+ age cohort by design (HOC 2019). There are over 1.3 million adults in the UK who are unbanked with no bank account and severe structural biases against those who have previously got into financial trouble through overdrafts (Age UK 2019). The Financial Conduct Authority has published research detailing how a third of unbanked people who have had a bank account before do not want one as they don't trust their own competence in reining in overspending (Financial Conduct Authority 2020[65]). There are still those who need help to open an account due to illiteracy. Those who would benefit from a basic checking account, say those in debt, are sometimes not offered one as bank's processes for determining what account to give someone relies excessively on credit checks. Older people with no means to prove their identity through permanent address documents in their name, or those who have led peripatetic lives and have moved often face significant barriers. Same applies to those without a passport, driving license or UK paper utility bills in their name who despite having a legal right to a basic bank account are still unable to access one.

Redesigning pension delivery systems through radical disintermediation that strips out layers of redundant or non-value-adding processes from service delivery does in fact drastically reduce public sector spending. However, by merging elements of tax systems and social security provisions and by relying on banking system's existing national infrastructure to deliver core pension payments, unsurmountable obstacles are placed in front of some older people reliant on pensions for income support. Eligibility calculation and identity verification[66] of someone who is over pension age is achieved by providing documentary evidence of the date of their most recent marriage/civil partnership or divorce, bank or building society details and the invitation code from the government letter about getting state pension as posted to their home address. These documents are to be submitted online on the gov.uk website.

When the simplified, automated version of verifying eligibility involves back-end processes that rely on matching data points held by different government agencies, this system is

---

[65] Age UK's written evidence to the Treasury Committee Inquiry on Consumers' Access to Financial Services makes the case that in an increasingly financialized economy, banks need to do more for older customers as they are still heavily reliant on free access to cash. Their submission demonstrates continuing age discrimination through a 'loyalty penalty' where older people who have had accounts over their lifetimes in a bank suddenly find their local branch shutting down and cash counters in remote parts of the UK subsequently disappearing. This has driven a reliance on unsecured credit especially in those retiring with unpaid mortgages. They strongly recommend that 'free access to cash is vital to many older people and must remain with excellent geographic coverage for the foreseeable future'.

[66] In late 2021, the Government Digital Service (GDS) offered estimates to the Parliament's Public Accounts Committee that 20% of the wider population do not have a form of identity such as a passport or driving licence making it harder for them to keep their commitment to serve all users as equally as possible. Representatives of GDS conceded that a lack of documentation or financial records makes it harder to make digital work for people on lower incomes.

optimised for those who are digitally savvy, have good credit checks, are in good standing with their bank[67], earned a good income over the life course, and have their names on key home ownership documents. In trying to personalise and create a do-it-yourself government, these data scientists, despite best intentions, don't create an 'intelligent centre design for the welfare state' (Margetts and Dunleavy 2013:9), but end up amplifying the vulnerabilities of the very demographic they are meant to serve. The latest ONS figures provide clues about the groups at most risk of digital exclusion – nearly 1.6 million women aged 75 and above live alone and a large portion of households without internet access are occupied by single female pensioners in that age bracket. There is research into how those with mobility problems are 1.44 times less likely to be using the internet than those without problems and individuals with memory problems are half as likely as those without to use the internet.

In nudging pensioners to complete their claims online by themselves, the task would indeed be less monotonous for frontline staff members and deliver significant cost savings. However, the quality of the citizen's experience in lodging a claim and managing direct payments depends entirely on the wider context of their life. This service transformation effort through digitalisation, 'joining up' and automation has the potential to exacerbate stress for a vulnerable older person facing multiple uncertainties in terms of their own longevity, ill health, difficulties in managing money, death of a partner and access to social networks such as someone to drive them to the bank. Navigating the invitation code, inputting it online on the correct gov.uk webpage and managing interactions with the financial institution that holds their nominated bank account can be a harrowing experience for those who are digitally excluded older people who despite being internet users may not be confident using technology for financial services.

If a datafied model trained on child safeguarding information creates a measurable type of a potential victim based on what people like you have historically done ('looks like an abuse victim'), relying on digital banking records and HMRC data points in real time favours a certain kind of pensioner. We saw in a previous chapter (Chapter 4) how abuse victims who live off the digital grid and care leavers whose historical records remain with another local authority district cannot have their journeys expressed into data and won't be picked up as a potential victim by the modelling process. Similarly, recombining HMRC and banking records into a datafied, joined up system 'sees' a certain way and it is implicitly built for the digitally competent older person with their names on relevant documents. I heard from interviewees

---

[67] That is someone who doesn't have gambling or large debts.

working on regex code how the matching sequence of simple characters like someone's name from marriage documents, address information and income declaration is not straight forward. Spelling mistakes, special characters in personal names (like, say, Gaelic accented words) or even something as simple as the inputting of a space into the wrong tab can generate errors on a large scale and can be stressful to an older person who is navigating this digital process by themself in a single person household at a vulnerable point in their lives. This kind of datafication is optimised for the digitally savvy, healthy pensioner who is heading into retirement having paid off mortgages and financially sound (without crippling debt or predatory debt collectors to fend off). These circumstances are not shared by the vast majority in these fraught times and especially not by those who rely on fortnightly pension payments to make ends meet or pay caregivers.

As we are reminded by Eubanks in her comprehensive ethnography of the automation of welfare provision and social security eligibility rules in the US context, people going through trauma or ill health need hope and need to know someone is paying attention (Eubanks 2018: 62). Even the most humane, benevolent, zero touch, choice architecture that nudges an older person to submit eligibility evidence online and employs checks and continuously updated income/national insurance contribution records to process payment can be a poor substitute for an actual human who can guide them through a significant life event. It is well established that digital competence and comfort level decreases with age. Especially if dealing with compounding stressful factors such as needing to be off the grid due to large previous banking overdrafts, illness, dementia and memory problems, anticipatory data-based decisions made even 'before you engage with us' (Interview 2) can create serious harm. The belief in objective quantification of a person's life circumstances through data held on them by the digital service draws on trust in the correctness and competency of 'institutional agents that collect, interpret, and share (meta)data culled from' databases maintained by the HMRC and the banking system's existing national branch infrastructure (Van Dijck 2014).

While the stated aim of this data driven vision is to increase 'join-up around significant life events so people only need to tell us about a bereavement once', life events are different for everyone and what data trails reveal are a necessarily partial view of the benefit claimant. The next section demonstrates how these data based 'customer insights' generate a proxy of the user that is inextricably entangled with the heteronormative assumptions coded in at design when the data on record is used to understand the real person and pre-emptively suggest if they are eligible for carer’s allowance or child benefits. The interviewee who rightfully points out that a lot of fraud is accidental is well intentioned; and that a future system could exist where the HMRC could just use the data it holds on him and not rely on the citizen to tell them that they have crossed a legal threshold and are not eligible for child benefits anymore.

The personalisation aspiration to understand the ‘real’ welfare recipient as a person and the friendly acknowledgement of fraud as sometimes being accidental that could be mitigated by full reliance on data points does signify an ideological shift. This does represent a softening of the strict conditionality regime where new technologies were being used to investigate and prevent fraud with the burden of proof shifting on to the individual welfare recipient. With the slow but sure retrenchment of the welfare state over the latter half of the last century, services moved from being universal and decommodified to a system where risks were being externalised through exposing job seekers and other vulnerable claimants to market logic. This meant increasing their obligation to demonstrate that they are meeting eligibility requirements and using benefit payments productively.

The onus was on claimants to demonstrate their deservingness, by making sure that their behaviour while on benefit complies with expectations of responsibility where they were logging in proof of job search activities, ensuring that their children were attending school while on child maintenance allowance and demonstrating that they were working with practitioners to accept help in tackling their problems. Under this welfare conditionality regime, the role of well-resourced benefit fraud investigation units were assumed to be crucial in recouping overpaid benefits, correcting problematic behaviours and enforcing severe sanctions. Therefore, the shift from the suspicion that everyone on benefit is prone to commit fraud which operates with a generalised presumption of guilt to the friendly acknowledgement that fraud can be accidental and that the department could be doing more to prevent it by using the records they already hold is significant. However, despite the rhetoric and stated ambitions, the implementation of this datafied infrastructure has continued to cause a lot of distress among vulnerable welfare recipients[68].

There is an ongoing push to continuously capture data from the service in a manner that makes it possible to know a population and individualised life events in a granular manner. Despite

---

[68] There have also been legal challenges through judicial reviews where the responsible Secretary of State was censured for operating an unlawful, unpublished policy. They were asked to reveal the algorithm that ‘targets’ the disabled for benefit fraud investigations and the data processes used to identify claimants who would later be issued notices to recover overpaid benefits.

lofty ideals, there hasn’t been enough consideration of the adverse effects of the system that they are building with so much care while operationalising the policy goal of making better use of data through the automation of this department’s verification processes and increasing the use of analytics. The next chapter draws on the above analysis to consider the role of proxies used to understand the ‘customer’ through measures for ‘customer closeness’ and how this process algorithmically groups and categorises claimants in insidious ways that don’t leave them with recourse to protection under the Equalities Act or current data protection law.

Chapter 6

6. 1 Proxies and Heteronormative data infrastructures: the price of ‘joined up’ welfare provision

We have seen a predominant aspiration to use data points to understand the claimant as a real person and plans to produce a ‘joined up’ experience where the department is being reconfigured to support people based on their life events. These include major life changes such as birth of a child needing the issuance of child maintenance allowance, unemployment benefits for the working age population and health assessments for those on personal independence payments and so forth.

> *As a government, if you know those two things, my national insurance number (NINO) and my bank details, you should be able to say; Yep we are paying you carers allowance or no, we're not paying you carers allowance. At the moment, this is a 30 page form!*
>
> *…that's the vision, which is that all of those [frontline] agents can be doing the real hard work stuff, of making sure that I know that I'm eligible to apply and making sure that I'm being supported in my caring work and in my caring journey… If my daughter comes off carers allowance and comes off PIP, what are the next steps for that? Getting her into education, getting her into a job.*
>
> *When that happens, what are the next steps for me? How do I get into a job? Because I haven’t been working for 10 years looking after my grandma or my daughter? At the moment [Department A] is falling short on its duty in those wings because it's pure focus is 'yes, we'll pay them and no we will not pay them, job's done.'*
>
> *Interviewee 2, Senior Delivery Manager, October 2020*

In seeking to get rid of the 30 page paper form where someone would have to fill in details that a frontline staff member would then verify, this manager hopes to shift to a ‘straight through’ online by default, disintermediated process where details held on the digital system are used instead. Replacing the current workflow would be an automated process where someone’s national insurance number (NINO) and bank details are used to generate prompts that talk about next steps for the male claimant who is living with, say, a daughter who is on personal independence payments and a wife who doesn’t work as well. The key advantage cited by this senior manager elaborating this vision is that of time savings as the current workflow has an overnight or three nights or six week processing delay based on the complexity of the claim. This is framed as a new data driven process that reorients the work from a transactional focus on payment to supporting the claimant’s caring journey by suggesting next steps using data.

The incoming Government Digital Service (GDS) head shares this view of putting 'mobile-first and hyper-personalisation' at heart of future service transformation (Markson 2022). He seeks to eliminate the need to send in information that the government already has on a citizen because it used to be too difficult for information systems in government to link up to another department and retrieve it. This service redesign using seamless data sharing is framed as simplifying the service for the benefit of the user and going beyond touching up the front end of the process or the user interface.

> *"For every good service that we have, we have a service that doesn't actually work on a mobile and hasn't been touched in 15 years – for good reasons – or a service that is really good on the front end, but you still need to go and find a printer and post it in," he said. "Or [there are some] services that are not really written in English – they are written in lawyer… [and] are almost impenetrable for users. Or there are services that ask you to send in information that the government already has on you, because it's too difficult for us to ask another department for that information. That's putting an extra burden on the system and the users."*
>
> *Except from interview in GovTech trade magazine, January 2022*

This means that at this pace, in about 5 to 10 years, this department will reorganise its data analytics and engineering efforts to capitalise on the wealth of data that is available to them and make services personalised around life events. Soon, self-declared details about oneself and household circumstances would not be required and the system will self-populate based on data on the citizen held in a different department so as to reduce the number of questions to answer on a form.

The technology used to collate these data points held on a citizen and understand the 'real person' so as to suggest the next step in their caring journey is through a set of Google Analytics tools that map someone's movement across webpages while they complete a form online. Using the example of a carers' allowance or a child maintenance payment that the interviewee brings up to illustrate his point, is it necessarily good to link a female carer's national insurance number (NINO) to her partner's and form a view of the household composition that the claimant has not voluntarily provided? Does this hyper personalised vision through data involve linking to the income declaration at the HMRC with real time combined income checks for operational purposes? In the name of delivering personalised data driven services, would it be reasonable to send someone in their 30s who is female, pre-emptive prompts about child tax credit expecting them to have a child? Whether or not these prompts come from a 'smart' data driven infrastructure based on inferences from data points held on majority of people on the system who would have children at that stage in the life cycle, this anticipates and raises serious questions about the datafied welfare model being built. At this stage, it becomes necessary to think through the consequences of the personalised nudging infrastructure being envisioned because of the potential to simultaneously build sophisticated data technologies that profile and exclude.

There is a growing field of research that documents the encoding of gendered expectations into data infrastructures and the irrefutable manner in which the UK welfare benefits system favours those who conform to gender stereotypes (D'ignazio and Klein 2020, Theilen et al. 2021). As will become obvious, welfare architecture has an antiquated view of the family unit as stable over time, women’s roles as carers and stereotypes about the ‘male breadwinner’ that haven’t really evolved over time. Carter (2021) argues that the reliance on these heteronormative assumptions mean that the benefit lines being digitised discriminate against those who do not conform to gender stereotypes. Child maintenance payments, nudges to lone single parents to return to work and universal credit which is based on a household means test paid into a single bank account all undermine women’s financial independence; and perpetuate the norm that ‘women should be supported by a male breadwinner, with state assistance only when such support is impossible to obtain’ (Carter 2021: 5).

Social Policy scholars have mapped the historical origins of the idea that institutions responsible for distributing welfare benefits during the post war period were averse to taking on responsibilities they believed were family responsibilities (Noble 2008, Hunter 2016). Welfare policies in the post war period were based on the expectation that men would work bringing in a family wage. Their female partners worked at home and were disincentivised from working by the structure of welfare benefits which focussed on family as a unit and assumed that the payments received by men would also benefit women and children (Fredman 1998). The dominance of this idea of the male provider continued through major reforms such as the introduction of social security against loss of wages (as part of the National Insurance Act 1946) which was mandatory for men and single women. Married women however were not included in this, and lost all credit from contributions made before their marriage. As Victoria Noble (2008) argues, since the assumption was that they would be supported by their husbands, they did not need state protection. Consequently, they received lower benefits and it took decades to recognise women as citizens in their own right within the welfare system. By the early 2010s there was recognition that caregiving roles within families were shifting as were women’s roles as economically independent individuals. The 1980 Social Security Act allowed for either partner to claim non-contributory benefits like child maintenance or disability benefits and from 1988 couples were able to choose which partner claimed these (Fredman 1998).

Despite not being explicitly articulated and valorised, the norm of the 'male breadwinner' persisted even as the family policy shifted in the 80s and 90s from being based on marriage to being based on parenthood and its associated responsibilities (Lewis 2000, Carter 2021). By the end of the 20th century, lone parents – 90 % of whom were women – were still being blamed for the crisis of the welfare state and it was assumed that returning them to work was the solution to child poverty. The rationale was that these single parents did not recognise the value of work and were state dependents by choice. This demonisation of women on welfare continues to this day as seen in media announcements at almost every recent Conservative Party conference; where there is always a promise to end welfare dependency and an exhortation to those who need social care assistance to rely on their families first before turning to the state for help (Javid 2021). In reality, however, rather than shirking the labour market, most single mothers do want to return to work but at a time that makes sense for them and under circumstances that make paid work viable (Grabham and Smith 2010). The introduction of Universal Credit as a landmark welfare policy reform which combines several individual benefits is significant in this regard because it has the most advanced use of data technologies; and being a brand new service built digitally from the beginning has embedded behavioural insights that can be aggregated to 'understand the real person' with uncanny precision. Universal Credit is paid into a single bank account based on the income of the household as a whole. For women partnered in co-habiting, heterosexual relationships this new regime means a return to financial dependence on men based on stereotypes about their familial contributions and social roles. Thus, Universal Credit as the latest cutting-edge policy product returns the female claimant to being recognised by the state through her male breadwinner.

Given this heteronormative data infrastructure (Hoffmann 2021) that systematically decentres women and single parents, data scientists designing 'slick' services personalised around life events are unlikely to think about the disadvantages of linking women to their male partner's NINO and bank account details[69]. There is a stated ambition to move towards a system that

---

[69] Given the discussion in the last few chapters, the obvious disadvantage is the possibility for coercive control through financial abuse which is not uncommon in dysfunctional and exploitative domestic relationships. There is also the question of how long the national insurance numbers and bank details should be linked in the case of couples separating without declaring it to the department or women escaping violent relationships without wanting their former partners to know where they are.

relies on the digital records available to the department and a drive to almost make decisions using data points before someone voluntarily engages with them through a welfare claim form. In section 5.2 we were introduced to the ‘test and learn’ logic of releasing three versions of a child maintenance payment application form into the live service where it became obvious that claimants were finding questions about childcare sharing patterns stressful to answer. In order to set up and maintain a child maintenance financial arrangement between separated/divorced parents there was a detailed application to be completed. To calculate how much money is due to be paid, an indication of whether or not the child stays at the other parent’s house was required. The data scientist in charge of the digitisation of the benefit line spoke about how among parents who share custody of children, sometimes, there were no set arrangements; while the question was set up in the format of a fixed choice questionnaire seeking details about the fortnightly arrangement (number of nights the child spends per parent[70]). Given the personal, intimate nature of the data on child share arrangements that the department now has, looking to the future, how much joined up information should the government hold on an individual, their household composition, and their relationships?

In the previous local government case study, it became evident that the team implementing the scoring system had to quickly come up with working rules about what data would not be shared between the police and the council. There was a broad agreement that it would be outside the remit of the council’s safeguarding responsibilities to provide the address of someone held on the council database to the police to issue arrest warrants or pursue an enforcement agenda. When all sorts of data fragments held on a couple can be linked in a digital infrastructure that has historically been conceptualised and designed to see women as an extension of their husbands and built on a cultural inability to value the economic autonomy and security of single mothers, a range of questions arise. In the name of a more caring, benevolent welfare state that responds with prompts for next steps even before you seek its help, would a personalised service based around life events for women of reproductive age involve prompts and suggestions about child care options? Would these be appropriate if these women were involuntarily childless or if they had been parents but had been bereaved? The allocative harm

---

[70] Following complaints from frontline staff and lots of calls to the call centre about the question they released a different version of the application that changed the wording and added free text input options. This kind of live iteration of the form was spoken about as using a/b/n tests and an experimental ethos to gather ‘user need’ in real time where they quickly found out that the current version didn’t work well at all.

of being misclassified by the automated suggestion when data tries to understand you as a person is very real. Would the onus then shift to the woman to report to the department when her life course deviates from the norm; and the support needs identified by a datafied system trained on majority behavioural data traces continues to not apply to her?

Crawford (2016) and Gillespie (2014) refer to this as conjuring a 'calculated public'. When everything is digital by default and the ambition to get claimants to send in information only once so as to create a gentle nudging infrastructure based on recorded data points is overlaid on to a heteronormative system, it invokes and claims an ideal benefit recipient 'with whom we are invited to feel an affinity'. Everyone else becomes a category error. That ideal type with an imagined birth, death, employment milestones and other life events may 'overlap with, be an inexact approximation of, or have nothing whatsoever to do with' the population that depends on support from this department (Gillespie 2014: 189). When data technologies are put in place to map in a fine grained way, household composition, dates when the female partner paused work for caring duties and child sharing arrangements, it is worth asking whose experiences and life events are erased by this welfare architecture. Respondents wouldn't answer questions about whether certain types of single parent households without a good credit rating are earmarked on the system as being at high risk of fraud and other targeted interventions. If there is an overarching ambition to use data provided on a form for one purpose to personalise the overall service so as to generate a 'straight through' digital experience, there are no formal guidelines in place to limit the secondary uses of data from the service – say for instance, using data on child movements between parents[71] in a risk model that calculates probability of fraud.

6.2 Path Dependencies: Policy intent and how it shapes data systems

It is no accident that a taxonomic system which erases those who don't conform to heteronormative stereotypes (now with data technologies that can pinpoint them in a precise manner) came together in its present shape. The ideational grounding for the Conservative government's family policy and drastic interventions into welfare reform in the shape of Universal Credit is provided by an echo chamber of think tanks such as the Centre for Social Justice founded by Iain Duncan Smith. In a series of agenda defining policy briefs

---

[71] As provided through a shared child maintenance financial arrangement form.

commissioned by the then Prime Minister David Cameron called 'Breakthrough Britain' (Centre for Social Justice 2007), they idealised families as the foundation of a healthy stable society and linked 'fractured families' and family breakdown with poverty, drug use, truancy, worklessness and all other social evils. There was an explicit intention to channel this loathing for lone parents and preference for 'two parent family formation' into policy design through tax credits and a system that incentivises couples to stay together; and goes so far as to come up with ways to get married couples tax breaks that lone parents don't. They called for an end to the Working Tax Credit's 'couple penalty' which meant that two-parent couples only received the same amount as lone parents which to them 'effectively discouraged two-parent family formation' (Centre for Social Justice 2013) and penalised marriage. Drawing on social conservatism with its focus on an idealised notion of a strong family which is always defined in heteronormative terms, a range of policy recommendations were made; the most outlandish of which were suggestions to prescribe relationship education and support lessons to prevent future family breakdown.

The idealised manner in which families as the foundation of a healthy stable society are conceptualised leaves no room to think about abusive relationships that need to dissolve and can't be fixed by relationship education lessons which would compound the harm to victims. This fixation with keeping the family unit stable as a precursor to stable societies comes at the expense of broader analysis of the complex gendered dynamics of violence, power and care within families. These functionalist ideas about the value of two parent families in reproducing stable societies resurged recently in the context of post COVID, cash strapped welfare state where demographic policy ideas were floated about a negative child benefit to find funds to look after social care costs of an ageing population (Morland 2022)[72]. This essentially meant taxing childless people for their failure to provide the next generation of productive citizens. Therefore, the welfare changes, tax credits and UC architecture designed with this ideological underpinning inscribes instrumentalist and utilitarian ideas about women's roles in caregiving,

---

[72] These policy interventions are meant to offset the continuing low birth rate and couples choosing to have fewer or no children. Other commentators have pointed out that problem is a structural one and focussing on individual reproductive choices get us no further - a decade of policies that prioritise the retired over the working-aged have created a society in which renters are insecure and home ownership remains a dream. Unaffordable childcare and rising cost of living costs have instead led to an inhospitable environment for raising children.

within households and society into the architecture of welfare, pensions and child maintenance policy.

Given this ideological baggage, it would be naïve to think about data points on lone parents, women and families held on the system as neutral (Broussard 2023). There is a stated policy intent to encourage two parent family formation through tax credits and disincentivise the supposed 'moral decay' of single parenthood or indeed an economically independent childless older woman. As categorisation systems whose existing architecture is being used to build datafied pre-emptive prompts around life events, they incorporate and retain the power biases of those who designed them and mutually reinforce these gendered expectations (Noble 2018).

Aggregation and data analytics techniques which are applied at a remove from the point of data collection does not ever appear as a direct intervention in the life or body of an individual person. However, creating a composite picture of a person from disparate datasets and digital proxies that may be innocuous on their own but produce deeply personal insights when combined (Metcalf and Crawford 2016) and overlaid with data captured over the lifespan. Legal scholars have begun to point out how algorithmic groupings where one is involuntarily assembled into a group based on clicking behaviour or web traffic is a form of proxy discrimination. Wachter (2022) posits that groups like 'dog owners, sad teens, video gamers, single parents, gamblers' are routinely assembled based on browsing behaviour, cookies from the webpage you landed from and one's online purchasing profile. Other such categories include single women with a disposable income who are courted by marketers as 'professional aunts, no kids' (PANKs) and low-income users assembled through data-based interpretations of an online user that are hard to definitively comprehend. Even logging into an account from a clean browser[73] does not fully wipe out one's online data trail or metadata signature. In an inferential economy where clicks matter, these increasingly incomprehensible groups 'defined by parameters that defy human understanding such as… clicking behaviour, electronic signals, or web traffic' are being used to allocate resources and reroute claimants (Wachter 2022: 1).

These decisions based on data signatures significantly impact lives of vulnerable people. This can range from referring someone for extra support because their 'backwards and forwards' movement in an online form looks like they are suicidal or in distress; or certain online

---

[73] It is worth noting that under the Regulation of Investigatory Powers Act (RIPA) this department is allowed to request the full browsing history of people they are investigating for benefit fraud.

activities look suspicious or fits the data template of risky behaviour rendering them subject to fraud investigations. Research to date has not investigated how these profiling technologies from marketing and ad placement systems are used in social care provision now that applications, commitments to meet conditionality requirements through logging in evidence of job searches and uploading additional documentation such as ‘fit notes’ is all online through gov.uk websites. Google Analytics 360 is the package that runs on the back end of the gov.uk platform and is used by performance analysts to understand user journeys and behaviour.

Wachter claims that the harms created by this kind of profiling and user segmentation into algorithmically generated groups that are inscrutable but do not map or correlate with legally protected groups such as sex, disability or race need our attention. It is very hard to tell how the groups computationally generated from browsing behaviour like ‘dog owners, single parents, gamblers’ and so on map onto the nine characteristics protected under Equality Act 2010 such as age, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, race, religion or belief, sex, and sexual orientation. This is especially important as the benefit claimants who are dependent on this department for income support face multiple complex vulnerabilities and submit personal data. Even after de-identification, it is possible to form a composite picture of a person compiling intimate details such as their childcare custody arrangements, illness, and financial information. Following Wachter’s provocation that there is a gap between legal doctrine and emergent forms of algorithmic discrimination where people are assembled into groups without their consent, it is necessary to check if there is a form of unintentional redlining being perpetuated by datafied processes. It would be long overdue to check using internal processes if datafied proxies for users built on service usage data overlap with these protected characteristics leading to different outcomes for them at a systemic level.

In addition to algorithmic groupings based on data trails and browsing behaviour, a key category that is not protected in the Equality Act is that of class and socio-economic status. Given the reliance on credit checks by third party agencies which is highly exposing of socio-economic status (Ada Lovelace Institute 2020[74]) and identity verification using bank details which requires one to be in good financial standing, protected characteristics need to be expanded in view of existing data practices. A comprehensive review of digital literacy and language capacities need to be conducted to understand if indeed backwards and forwards

---

[74] The Ada Lovelace Institute suggests that given the current state of consumer profiling, even if the public sector agency were to conduct an equality Impact Assessment and make it public, it would be insufficient in the various forms of discrimination that automated decision making systems perpetuate.

movements in the linear screens of a form signifies a problem or suspicious behaviour. It could just be the browser getting stuck as it is being used on an old computer or someone with low literacy skills navigating the system.

> *So we don't track… We don't track mouse movements but... So sometimes we can… depending on the sorts of questions that are asked in the journey…we can learn things about the users. But we do it in an aggregated, anonymous way.*
>
> *[following question about whether she could elaborate how]*
>
> ***So I think, predominantly in digital analysis, we look for those common themes and like the majority behaviour.** So if we see lots of people displaying that backwards and forwards behaviour, then we think there's probably a problem there. If it was only one or two people, we probably wouldn't consider it to be a problem. And because we tried to get a really good cohort of users and say, we tried to get people with accessibility needs and a really good spectrum of the sorts of people that use a service that helps us to identify potential problems that are happening on the site... I think there's always… it's always a thing about prioritisation. So it could be that there is a problem that only affects one or two people. But it would end up being prioritised below something that was suspected to affect 50% of the people that use the form. So we do, we do find some issues, but we have to prioritise based on what is going to make the biggest difference to the biggest number of people. Unless, that issue means that someone absolutely can't apply for that benefit. But we also… we have offline routes, for a lot of our benefits.*
>
> *Interviewee 12, Senior Manager – Data Science group, July 2021*

The manager in the interview makes a strong case for ensuring their digital service works for the majority of the population through user insights drawn from the service in an ‘aggregated and anonymous’ manner. We have seen how there is an overarching ambition to match data points held by different government agencies; rhetorical aspirations which still remain aspirational for some parts of the organisation. When there is a demonstrably heteronormative undercurrent to the architecture of the welfare bureaucracy whereby it is possible that in the name of life event based service optimisation, pre-emptive suggestions about childcare options are made to an involuntarily childless woman, these assurances about anonymisation might not be enough to inspire confidence. We have seen how deeply personal information can be aggregated. Government has to be for everyone and not just a majority and in accelerating digitalisation to optimise based on majority data, it is worth carefully considering whose life cycle milestones are misrecognised by the system by design. Even if the design of the system and the datafied inferences that are drawn from the service are a problem to one or two people (Interview 12), it is still a massive burden for those who are locked out. The interviewee does admit that even if it does not affect a large number of people, if it means that those who are locked out can't apply for the benefit at all, they would prioritise it. The question remains whether there are tools in place to understand and include those who are being unintentionally excluded and misclassified.

6.3 'Edge cases' and category errors

In repeated interviews 'edge case' is a construct that was used to describe everything outside majority behaviour. These are categories in the design process where users with niche needs are talked about as 'edge cases'. An edge case is a problem or situation that occurs only at an extreme (maximum or minimum) operating parameter. In programming, an edge case typically involves input values that require special handling in an algorithm. According to the GDS assessment framework they operate within, the goals are broadly about making sure that most users can go through the digital 'straight through' process easily. So the focus tends to be on the bigger groups of people rather than the edge cases.

A senior delivery manager in charge of shaping the work priorities of around 300 staff members explained how in digitalisation, the human frontline processor's intuitive sense of whether to believe the paper evidence in front of him has to be translated into a series of programmatic steps to put into the computer. The experiential wisdom that makes him think a detail is worth investigating is referred to as 'fuzzy logic' (Interview 2), that is hard to put into code in a way that applies to all claimants equally. When an experienced human processor decides something doesn't look just right and needs to be checked against records on file they are acting on practice wisdom that took years to learn. For example, considering the specific circumstances of the COVID lockdown in 2020, this would apply if someone self employed as a plumber making £6000 in the previous year were to file for carer's allowance while caring for his sick daughter. This plumber was not able to earn anything in the current financial year and without the benefit could not put food on the table. He is in a complicated situation in that it is his girlfriend's daughter that he is caring for and as a self-employed man HMRC holds data on him that isn't optimised to link easily with this department's income verification systems. This application was described as an edge case as it is a complicated, tricky set of circumstances that might cause the frontline processor to take some time to think about and double check some of these connections as the claimant was able to earn a living in the previous year but not at the time of the application.

> *So that's an edge case; where a person would look at it and go, 'Oh, this is so complicated'. It is going to take some time to think about and just double check all of these connections. And they're not married, is that relevant? Is he dependent? Maybe his daughter spent 12 weeks in hospital in March, April, May. But is now living back with them?... what's going on there? So it's a... it's a complex case.*
>
> *Interviewee 2, Senior Delivery Manager, October 2020*

Like all other government departments, at the time of the interview, this group was also having to change its processes to comply with the changes introduced by Brexit. In processing claims from UK residents with a ‘foreign sounding name’, would they have to retrieve data from the passport office to ratify the claim of having a British passport? The legislation wasn’t clear on these aspects and decisions had to be made whether to code in checks for all passports or none at all. There are series of checks that need to be carried out in an ‘edge case’ scenario; some of which would be going beyond the requirement of the law. When a human processor handling a carer’s allowance application makes a decision to check the passport details of the claimant seeking support as they claim their wife is also out of work, he might choose to check their NINO, their wife's HMRC records and their marriage certificate. This is most probably based on previous experience that raises suspicion about the legitimacy of the application and because at some point in the processor’s career, during the check they discovered something new.

However in the digitalisation process that displaces the human processor, decisions had to be made about what checks are required by the legislation and programme them into the computer. Given the live iteration and ‘test and learn’ ethos that was initially encouraged, most of the ‘complicated’ cases or claimant circumstances and needs that could not be catered for after digital transformation were discovered as category errors in the backend of the digital service. More often than not, these complicated needs were brought to their attention by advocacy groups like Citizen’s Advice or Child Poverty Action Group raising concern or from internal Serious Case Panels.

In deliberatively designing in this manner, some people become edge cases when their problems are not shared by the vast majority of claimants who find the digital service usable. Their circumstances and their unique yet deeply felt needs become occluded, othered and assigned a lower priority. In not critically reflecting on who the system is optimised for and thinking of the edge cases as technical problems to be solved once the majority accessibility needs are met, they are erasing (or at best inconveniencing) those for whom the ‘straight through' experience does not work. In the use of proxies and the new taxonomic systems that are being built through datafication, they are compounding the disadvantage of the real person behind the 'edge case'.

6.4 Conclusion

In this chapter, we saw how in converting social categories to data categories, some people's circumstances and lived experiences were relegated to 'edge cases' or rendered invisible by design. Despite characterising themselves as 'custodians of the customer experience' transforming the way in which people interact with the digital service by building better 'user experiences' through insight-driven strategies, data scientists and tech evangelists miss a key principle of welfare administration and policy making.

Policy doesn't have users and legislation doesn't have customers. Welfare and social care provision is not a website optimisation problem. Instead of users, there are a vast range of stakeholders often with conflicting interests. In trying to discipline the welfare benefit claimant into submitting applications and demonstrating compliance to conditionality regimes on a website, the datafication of social service delivery holds just one stakeholder in view which is the end user of a computer system.

> *In policy development, it's not 60 or 80% of the easy cases that matters. It's the 10 to 15% of extremely weird edge cases that matter, because you're creating law that applies to everyone in the country. And it has to be administered equally for everybody in the country.*
>
> *Interview 18, Former Senior Civil Servant and advisor to the Head of the IT Profession, Sept 2021*

In order to understand the shape of the data infrastructure in this case study and the reason for the extensive reliance on proxies, it is necessary to historicise the development of datafied processes. The initial establishment of Government Digital Services (GDS) and the political imperative under successive governments to cut welfare budgets to deliver on election promises created path dependencies, which meant that services were cut to the bone and no measurable type or prototype captures the full range of constituencies that the welfare administration system is meant to serve.

Government Digital Services as a group that was introduced at the apex of the central government in 2014 operated with the explicit principle that everything was being done wrong in government. They maintained a disruptive culture and made a virtue of purposefully not integrating with the rest of the civil service. The existing IT profession was rubbished as old fashioned, unnecessary and not fitting the modern digital era anymore. Established in around 2005 as a loose conglomeration of cross departmental staff with skills to run the big databases of that decade, the IT profession thinned out following tussles with GDS. GDS maintained that the age of the Chief Information Officer was dead and there needs to be a new way of doing government through web based systems and an agile project management methodology.

This was a moment in time that coincided with the proliferation of advice about the advantages of privatisation. A series of Cabinet Office driven policy reformulation attempts led to the idea that citizens expected government to be like Amazon and work with the dynamism of the wholesale retail model. ‘Direct government’ as a digital one gained policy salience with a plethora of private consultants being brought into to reform government by building more websites, and putting everything on the website as a service. Simultaneously, there was a long standing attempt to overhaul the welfare system to ‘make work pay’[75].

---

[75] Much ink has been spilled in analysing how the argumentation around the so called ‘poverty trap’ and the idea that the benefits system itself disincentivised people from work gained currency and legitimacy in policy communities from the late 2000s. The overhaul of the welfare system to ‘make work pay’ (Figure 6) was initiated on the advice of private consultancies that made the argument that for those dependent on income support, moving into paid work unleashed marginal tax rates and the swift withdrawal of benefits. The combined impact of taxes and the overlapping withdrawal of multiple benefits meant that many at lower earnings, faced a participation rate of 75% and were able to keep just 25p of each £1 they earned. It was argued that this made it easy for an imagined welfare recipient to just stay on benefits and never enter or progress into paid work (best illustrated in the rather patronising cartoon in a key policy document advocating for a single working age benefit that makes it better to work rather than stay on benefits). With this spurious reasoning began the most significant changes to the British welfare state since its inception in the 1940s and the fraught attempts to streamline benefits in a complex policy landscape where a central government department administered the out-of-work benefits, HMRC the in-work tax credits and various local authorities across the UK were responsible for delivering housing benefits.

Figure 7: Make Work Pay dilemma

Streamlining benefits was seized as one of the first digital exemplars that other departments that were starting their digitalisation journey were expected to emulate. Senior civil servants identified how the task of streamlining and digitising the disability benefits system and an overall cap on benefits paid to working age claimants was used as a way for Cabinet and Treasury officials to gain control over the social security agenda and the purse strings of the department implementing welfare benefits. The then minister at the centre of efforts to drive welfare reform and redesign the benefits structure recounts how political deals to cut the welfare budget were made in order to have something to brief the press on when the government of the day was under scrutiny on other matters (Freud 2021). Institutional suspicion felt by the Treasury for the social security agenda was combined with the need to demonstrate annual budget savings and prepare for the next election cycle. In this context where blaming welfare claimants remained politically expedient, cutting welfare budgets through a new streamlined digital benefits system became enduringly popular.

GDS therefore forced the website transactional model on Department A almost immediately after it was launched as part of the Cabinet Office. Self service by benefit claimants on a website was the goal which meant that there was an essential conflict of policy objectives on many levels[76]. Social landlords who used to be paid housing benefits directly by government were driven out of the market effectively reducing the supply of accommodation for benefit recipients and increasing homelessness. Housing allowance was rolled into Universal Credit such that social tenants would need to pay their landlords directly, reversing decades of policy where the Local Authority would make the payment securely and regularly on their behalf. This threw housing associations into financial crisis with welfare recipients building up arrears and taking months to get used to paying rent own their own. The 380 devolved local authorities across the UK who were responsible for delivering housing benefits were not consulted for two whole years. Agile as a methodology only works when all the key stakeholders in the system can literally be in a room at the same time and by prioritising building the website, they started at the wrong end. As several interviewees at Department A reiterated while providing background context, 'Cabinet expects Agile'. With the political objective to remove the 'barriers to work' and the expectation that this department was to be a digital exemplar, technical difficulties in setting up the digital infrastructure for the full Universal Credit functional service were rushed over. Despite tax credits and housing being the main structural changes this system was built on, the minister in charge reports being very worried about the accuracy of the real time income PAYE information being fed to them by the HMRC even as late as when it was about to go live (Freud 2021: 241). Instead of testing every aspect of the system before scaling up, the overarching concern was to have demonstrable changes before the 2015 general election (200,000 people on Universal Credit) to deliver on political promises. Frontline staff are crucial components in a welfare to work structure – I heard how job centre managers initially thought the digitalisation of the department would mean updating their IT systems, making it easier for them to have access to local labour market vacancies in a coordinated, real-time manner so that they can pass on that information to jobseekers and welfare claimants. Instead, a centralised website based architecture was developed by GDS and

---

[76] Much of this 'behind the scenes' history of the government digital profession and the changes wrought by GDS were pointed out by my expert Interviewee 18 who worked as an advisor to the head of the Digital Profession and later as an academic consultant for the EU. He continues to provide written evidence for parliamentary and the National Audit Office (NAO) inquiries into the digital design of Universal Credit. This is also covered by trade literature such as Computer Weekly articles following senior IT personnel changes and the 'Civil Service World' magazine online archive that I compiled as the documentary basis for this research. Autobiographical books by politicians and ministers active in government during this time also confirm these stories.

foisted on Department A which was all about the user experience and taking the administrators out of the loop and now putting it online for the end user[77].

The heritage of digital was thus forced on my central government case in a top down manner leaving them with centralised performance indicators and limited leeway when it comes to what they can do with data tools. Senior data scientists had to then redesign the data profession for the department so as to undo the earlier ‘cookie cutter’ (Interviewee 11) approach introduced by GDS. They are now conscious of how limited data from the digital service is in measuring things that matter to their sector mainly due to the set metrics around the rate of digital uptake that GDS has mandated through the service standard.

Especially at the middle management level, the tech optimism attached to data science has very clearly mellowed over the last four years. Organisational fixes included thinking of data as a product and recalibrating the role of data professionals to resemble that of digital professionals and echoing their methodologies. Therefore, data scientists are now stationed with teams that are digitising existing benefit lines to have broader conversations about measurement. This is a conscious change to move from a mode of working where an online form would be created that is an exact copy of the offline form replicating all its problems online; and then expecting data science to magically improve the service. The role of the data scientist has evolved to negotiate and communicate with digital product managers about what can and can’t be done and help them think about the digitisation process in terms of whether these are the right changes to be making. They are also starting to systematically benchmark whether the change made the difference that they said that it would do and thinking about whether digitising a services is being done in a way that improves the service overall through attempts to continuously collaborate with policy and operational colleagues who are closer to the frontline issues.

This however remains an ongoing struggle. We see how the aggregative power of datafication to assemble citizens into groups through proxies making inferences about backwards and forwards movements from Google Analytics is amplified when there is political pressure to

---

[77] There are well known stories of the Chief Information Officer at Department A resigning from the civil service in a fury due to frictions with GDS (Interview 18) and anger at being ‘reduced to being the post-boy for youngsters (GDS) who don’t have a clue’ (Freud 2021: 239). GDS maintained its image as ‘youngsters in jeans’ deliberatively disrupting stuffy civil service values – with the political imperative to deliver welfare reform digitally, they seemed to have found their groove in the lead up to 2015 after which their role transitioned into one that is less confrontational.

push through welfare reform. Path dependencies introduced by austerity measures in the last ten years and other politically driven imperatives have shaped the current data architecture of this department in profound ways.

Chapter 7

Justificatory discourses of Datafication: A 'test and learn' approach to Evidence Based Policymaking

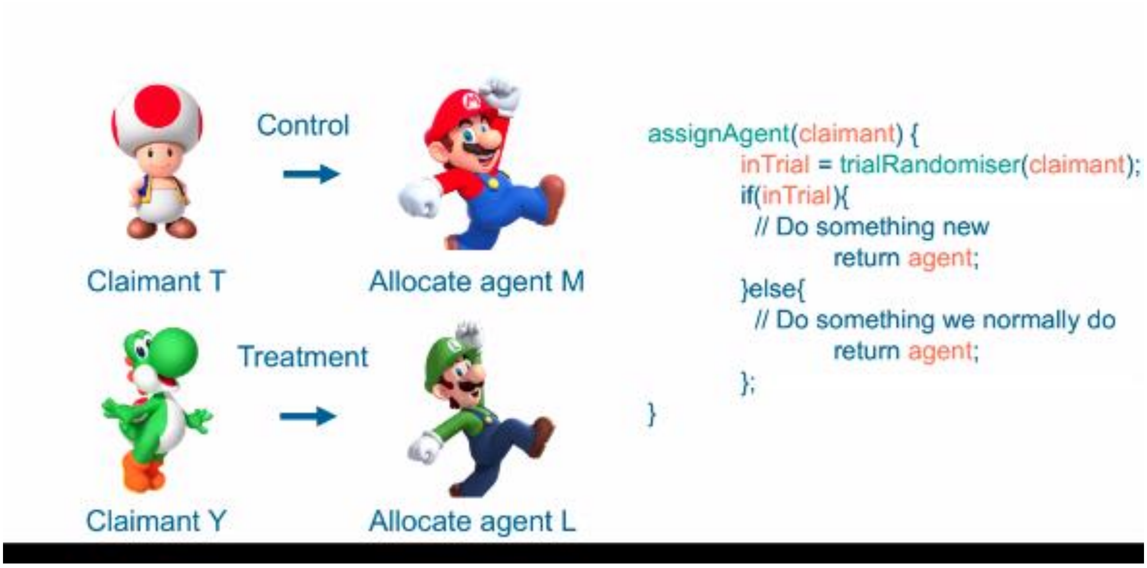
> *Interviewee: ...the reason why we exist is because in my opinion because if it didn't exist you would literally be throwing money away. Because, in the absence of... largely what we are doing is we are examining… we are creating evidence, right?*
>
> *You are wasting money if you don't use evidence. You're essentially spending money on something that isn't doing what you want. That's weird! It's not evidence for the love of evidence. It is evidence for the love of saving money. So, the real challenge that government has is to create* ***a culture of producing robust evidence and things like A/B testing are important components of that.*** *And then having people expect to see that kind of good evidence and people having the data literacy to understand it and make their decisions based on it… and I guess where are we now? Who knows? I think we're pretty good in [Department A].*
>
> *Interviewee 1, Senior Data Scientist, Sept 2020*

The introduction of predictive modelling, data science methods and advanced analytics enables the use of novel, computational ways of defining social risk in safeguarding situations and in the provision of welfare services. As emerging technological processes, their implementation is accompanied by various justificatory discourses that promote their use and strategically narrate and frame their benefits. The above data scientist advocating for the intensive use of automated A/B testing makes his case by appealing to a culture of expertise centred on evidence.

As we saw in a previous chapter, A/B testing refers to a technique where claimants of working age benefits are randomly split into large treatment and control groups that receive different variations of a feature of the digital service. In referring to A/B tests as the gold standard for measuring impact and as an exemplar of robust evidence, the interviewee is also affirming a new form of regulatory knowledge. He is claiming this large-scale user testing regime as a new kind of evidence for policy.

The data scientist introducing their A/B testing prototype considers the allocation logic in Figure 8 to be an entirely technical problem and justifies the trial which involves frontline staff providing services to vulnerable jobseekers by comparing its design to Random Controlled Trials (RCT) used in clinical medicine. It could be that as a result of being sent to Agent L, the jobseeker received bad advice or a rushed session when compared to the frontline agent that they would have normally approached. By interfering with the claims process in order to procure evidence, some unwitting claimants are exposed to sub-optimal outcomes (or inconveniences at the very least) that breach their capacity for autonomy and self-determination. These user experiments are nothing like Random Controlled Trials (RCTs) which have an established research protocol, clear ethical parameters, guidelines and are subject to independent oversight mechanisms. Furthermore, researchers conducting RCTs are required to register them before the trial begins and measure and document their unintended adverse impacts on human subjects.

Figure 8: Automated A/B test to hypothetically route a claimant to different frontline staff in the control group and treatment group



```
assignAgent(claimant) {
        inTrial = trialRandomiser(claimant);
        if(inTrial){
          // Do something new
                return agent;
        }else{
          // Do something we normally do
                return agent;
        };
}
```

Source: From slide deck of presentation attended during participant observation, Sept 2020

These interventions into the benefit claim process are aimed at finding ways to increase the efficiencies of the service; which several interviewees claimed were imperative as the same number of frontline staff have to support a larger number of payments over time. Interviewees repeatedly described how they had a mandate to run these trials to measure how long it would take someone to perform various tasks and compare options that require the least amount of time from their frontline staff.

> *And all we would be interested in is making, (pause) …if somebody does something, does that make time go up or down. It's not about predicting whether time will go up and down or predicting how much time it will take in the future. It is literally like at this moment in time, as a result of the change that we just made did the time go up or down?*
>
> *Interviewee, Data Scientist, Sept 2020*

This kind of disproportionate emphasis on measurement draws on an older discourse of public sector modernisation through audits and performance management. This earlier audit regime saw the ascendency of policy evaluation as a discipline and its formal institutionalisation; with explicit assessment standards routinely being built into programme designs and budgets (Solesbury 2002, Boaz et al. 2008). While this has always been ‘ex-post’ evaluation to see if a project has met its objectives, datafication has generated measurement techniques and tools of experimentation that make continuous, real-time evaluation a possibility. In contrast to previous analogue modes of record keeping and filing in the administration of welfare benefits, data from the digital service logs of Universal Credit are amenable to filtering, sorting and segmentation into a pipeline of these sorts of tests with relative ease. This amplifies an established logic of quantification and measurement so much so that responses to changes to the service are made knowable in an immediate, dynamic and fine grained way.

Therefore, the gradual normalisation of datafication is accompanied by a new kind of empiricism that involves constantly checking if something is ‘doing what you want’ through iteration and the extraction of real time information on how users are responding to the digital feature; and how design choices and settings are faring. Tools of datafication such as user experimentation, a/b testing, live iteration and new kinds of evidence that were traditionally never seen as the basis for policy-making are presented as superior, modern and more scientific than relying just on the discretionary capacities of frontline civil servants. This chapter thus will delineate the main justificatory discourses that frame complex policy problems as technical problems and elevate novel, uncertain data science methods to a credible body of knowledge.

7.1 ‘What Works’ and the discourse of Evidence Based Policymaking

In order to understand the legitimisation of novel data science methods and the instrumentalist view of data as evidence, it is necessary to look closely at the rise of the evidence-based movement in public policy. A prescriptive, rational model of planning and decision making promoted at the turn of the century, there was an overwhelming call to centre evidence in a bid to standardise and professionalise public administration. This drive towards evidence-based policy making gained renewed currency as the then Labour government set out its modernisation agenda (Cabinet Office 1999). It was hailed as signalling a new mode of policy making underpinned by rigorous and relevant scientific knowledge and offering great promise by synthesising evidence on the types of government interventions that are effective at addressing social problems[78].

Evidence in this context came to mean evidence of whether policy projects had achieved their intended outcomes and effects while delivering ‘value for money’. In other words, assessing effectiveness and accumulating evidence to that effect while providing information on performance against targets became a core aspect of the policy making process. With the increased commitment to a search for ‘what works’ (Powell 1999) and a results-oriented decision making process, evidence came to take on a narrow, utilitarian meaning. There is now a large body of research that shows how this emphasis created an instrumental rationality which conceives of policy making as a technical endeavour of problem solving aimed at identifying good means to achieve fixed, unambiguous ends (Sanderson 2002, Solesbury 2002, Gray et al. 2009, Kay 2011, Parkhurst 2017). Such a reductive view erases the normative basis of policy practice because policy goals are often contested and require the ability to question whether the end is worth getting to. Given the complexity of social problems that governmental attention is focused on, the aspiration to find the right ways to generate and apply knowledge of what does and does not work leads to increasingly sophisticated techniques of measurement; at the expense of restricting broader deliberation on whether we should be pursuing a particular policy initiative in the first place. From this perspective, the pursuit of ‘what works’ or what achieves given policy ends more effectively is not a purely technical task but rather one that produces a preference for measurability where social problems that are conducive to measurement are given policy priority as opposed to those which are important to affected populations (Barnes and Parkhurst 2014).

This instrumental rationality is clearly echoed in a similarly distorted view presented by interviewees where policy making and professional practice is seen as ‘context-free technical problem solving’ through specialised expertise and scientific evidence (Sanderson 2006). The following articulation of the need to embed data science into *Department A* draws on the key

---

[78] These ideas gained salience in the 1990s and were accompanied over the next few decades by the establishment of formal governmental and non-governmental agencies that work to increase the use of evidence in policymaking. These meant increased status of analysts within government and the establishment of strategic teams like the Performance and Innovation Unit (PIU) which were created to generate analytically driven solutions to cross cutting, long term issues within government (Parsons 2017). UK’s Alliance for Useful Evidence network lobbied for the creation of ‘What Works’ centres explicitly designed around the health model (Parkhurst 2017).

argumentative threads of the discourse of Evidence Based policymaking by preserving a hierarchy of evidence and discrediting other ways of knowing while activating notions of apolitical change.

> *Interviewee: If you don't have any evidence, then you are just doing things randomly based on personal judgement. History in any business, and the civil service included has told us that it is a very poor way to run an organisation. And so, in a large organisation like ours and in any others, there's always gonna be people who think they know what the right thing to do is.*
>
> *AT: Is that based on how they've always done stuff?*
>
> *Interviewee: Yeah, based on how they've always done stuff, based on prejudice, based on stereotypes, based on not enough information. And evidence is the counter argument to that - to help drive us down the right path ...If you listen to Michael Gove, if you listen to Dominic Cummings, Boris Johnson… They have all talked about the role that evidence has to play in the running of government. This is something that is expected… It is expected in the civil service that decisions that departments make are based on evidence. And that will become more of a specific requirement as time goes on, I think.*
>
> *Interviewee 1, Senior Data Scientist, Sept 2020*

By imagining a paradigm shift where previous modes of decision making in welfare policy provision were faulty and arbitrary, these data science methods promoted by the interviewee are discursively cast as the only methods that can measure 'what works' and provide a higher order of evidence. Personal judgement, institutional routines, organisational norms, informal rules, tacit values and other forms of experiential knowledge are dismissed in the articulation of this reformist zeal. Moreover, these are seen as contaminating the decision-making process. This interviewee goes on to talk about the much maligned 'wierdos and misfits' blog[79] post (Cummings 2020) authored by former senior policy advisor Dominic Cummings where he sets out the idealised set of skills for high performing policy teams working for the office of the Prime Minister at the apex of government. He imagines and frames computational rationality as being able to usher in a transformation within existing civil service culture through data

---

[79] This is a blog post that details desired characteristics for the policy advisor roles that Cummings was hiring for – starting with data scientists and an assortment of 'wierdos and misfits', he sought to rewire government institutions using data, modern cognitive technologies and dynamic tools that help understand complex systems and 'argue with evidence'. Some of these reflections on using technology to improve government decision making draw on ideas from Physics and IT to talk about the need for 'Seeing Rooms' for decision-makers; that is rooms designed to support decisions in complex environments. He compares normal Cabinet rooms, such as that used in summer 1914 or October 1962 (during the Cuban Missile Crisis), with state-of-the-art technologically augmented 'Seeing Rooms' that alter the physical architecture of Westminster offices to support a new way of working that understands problems as complex systems that can be tackled with data.

scientists with the capacity to translate ideas from computational physics, data visualisation and cognitive technologies and apply them to practical public policy problems. This problem definition of the civil service being rampant with prejudice and stereotypes; with *evidence* needing to come in to eradicate these messy, flawed ways of working invokes a ‘value free’, morally and politically neutral, technical objectivity that works as a legitimation strategy.

Evidence takes many forms. Critical social policy scholars responding to the Evidence Based policymaking movement have repeatedly shown how such an instrumentalist and scientised narrowing of the evidence base would marginalise other forms of expertise. Practice wisdom of experienced professionals such as front-line case workers and social workers as well as the organisational knowledge associated with managing program implementation would be types of policy relevant knowledge that would be implicitly suppressed (Webb 2001, Head 2008). In addition to the contributions of the experienced practitioner, the users of the service or the welfare recipients themselves bring a perspective on aspects of the service that ‘make a difference to their lives’ and this type of experiential knowledge would be devalued by the kind of formal evidence orientation conceptualised above (Beresford 2001, Lewis 2001). Seeking to root out value laden, opinion-based ideas through evidence erodes professional judgement and misrecognises the organisational and institutional context in which decisions are made. Therefore, in assuming a direct jump from evidence to a good decision without any space for the interpretation of evidence, deliberation and weighing up of different alternatives, practitioners are conceptualised as information processing systems whose work can be systematized, optimised, and managed. Although there is some recognition of this amongst the data scientists I spoke to who insist that human discretion is always in the loop (see chapter 7 for details), the dominant discourse draws on all these technocratic argumentative threads and this rationalistic fallacy retains a strong grip on prevailing ideas about how to improve organisational practice and policymaking.

Thus, we see the emergence of a historically situated public sector modernisation discourse that views intractable, multifaceted and long-standing governance issues as discreet, measurable puzzles that can be solved using technical expertise and more direct and instrumental use of evidence. Such a normative commitment to Evidence Based policymaking maps well onto the New Public Management preoccupation with performance management, standardisation, and cost effectiveness though the technologies of indicators, targets, and audits. Combining this renewed optimism with the managerialist preoccupation with measurable outcomes creates a potent justification for reinventing government using a new kind of evidence for policy. The data scientist quoted at the start of the chapter exemplifies this when he eloquently frames the effort of establishing data science as a field of expertise and transporting its disruptive methods into a largely bureaucratic department as generating rigorous ‘evidence for the love of saving money’.

From interviewee accounts, it becomes clear that majority of data science projects in operation that were reviewed as part of fieldwork seek to deliver on the rationalist dream of continuous evaluation and knowing demonstrable impact. No longer is it sufficient to systematically evaluate a policy initiative at the end of each financial year or after the piloting phase. By setting up avenues to gather responses to changes made in real time, a mode of continuous evaluation and performance management is activated. Economistic rationales and the claim to produce the best kind of evidence are used to legitimise live user experimentation and other sophisticated data scoring technologies described in the following chapters. The impetus to artificially create these trials, partition tens of thousands of claimants into control and treatment groups where different versions of the digital service feature are shown to different people comes from this need to demonstrate results and outcomes. It strains the administrative law principle of needing to treat everyone equally when implementing government policy. By treating welfare administration to vulnerable claimants as a website optimisation problem, this brand of datafied instrumentalism elevates quantification, precision and measurability over equitability and other principles of good governance.

These technological transformations are narrated as straightforward steps towards greater effectiveness and efficiency. However, it is worth noting that reconfiguring the logic of care in a manner that represents an expansion of the ability to monitor and regiment frontline professional practice has long term consequences (Matias 2017, Veale and Brass 2019). For example, some things are easier to measure than others. While it is easy to measure time spent on each task with unprecedented precision, in the case of the A/B testing prototype described above, it is harder to get evidence on why it took longer for the job coach to work with a jobseeker (perhaps the jobseeker had a disability that made it take longer to communicate their needs). Therefore, a system optimised solely by metrics around time, can prioritise some needs over others and privilege efficiency over a range of other public values.

7.2 Defining the value of Data in the Public Sector: From 'Silos' to producing 'one version of the truth'

The use of evidence in policy making has a long lineage and we have seen how the discourse of Evidence Based Policymaking is mobilised to justify new kinds of techniques of measurement with significant social consequences. If the claim to an older notion of evidence which is the ultimate pinnacle of 'scientific' policy making is one strategy, another way of legitimising data science expertise within these sensitive social policy fields is by distancing themselves from the known abuses of these techniques.

> *"…It is not like deep learning and stuff that companies like Facebook and Google are doing…. A company like Facebook is governed by their own self interests. Because we are bound by law, by policy and by ministers and things like that. So, I would be more concerned by bias in machine learning, which I think is hundred percent a problem. But I don't think the public sector is where the problem is manifesting itself."*
>
> *Interviewee 1, Senior Data Scientist, Sept 2020*

There is a very clear sense that, unlike the deployment of algorithmic systems and data science methods by companies like Google and Facebook that are driven by profit rationales, public sector uses of these systems deserve our trust. Interviewees with a managerial component to their roles explained how data scientists are deliberately chosen from the private sector to innovate and shake up the working culture of the public sector organisation. These normative distinctions between the private and public sector uses of data science are hard to sustain. The tensions in this emerging discourse are evident in how the data scientist insists that the way he is mobilising these novel techniques is markedly distinct to the worrisome ways in which Facebook and other private tech companies do.

If the use of data science in the public sector is said to be fundamentally different to the private sector, there seemed to be a consensus amongst interviewees that the existing utilisation of data and government IT systems were very unsatisfactory, outdated, and steeped in ways of working that were unresponsive, rigid and bureaucratic. There is a very specific vision of change that data science and the various methods and techniques grouped under the label 'data-driven practices' are meant to deliver. Several interviewees seem to be framing the rationale for their projects in response to the spectre of bureaucratic 'silos' that hinder timely and direct access to information and administrative records. Bits of data sit in pockets of the department making it cumbersome for another department in the same larger public sector organisation to retrieve it.

The frustrations of an external consultant[80] who was brought in to develop analytics models in order to recommend service improvement changes at a local government authority are evident in the quote below.

After expressing his shock at the differences in the organisational culture of the government department he was consulting with, he sounds incredulous that it is impossible to identify something so basic as how many households in the borough were known to both the Children's Social Care and Adult Social Care teams at the Local authority due to the siloed database design structure.

> *Interviewee: ...there was often no unique, no common unique identifier, between databases. So that meant when we were trying to build up like a common picture of a single individual, that was often deemed impossible to do. Or even if we were trying to identify how many of the same households were known to Children's Social Care and Adult Social Care. And that was also deemed, like, not a possible query to answer by local authorities… Yeah, in terms of like, the, the systems that we use... the case management systems that were used, there would be one for Children’s Social Care, and one for Adult Social Care. So getting them to talk across was very, very difficult. And my experience was that the data quality in local authorities was really poor. So what I would have thought of as simple queries to answer, weren’t [simple or feasible].*
>
> *AT: What are some of these queries? What were some of the questions you were trying to answer using data?*
>
> *Interviewee: Yeah. So the one that sticks out… I remember asking how many of the households in the borough, were known to Children's social care and Adult Social Care? But they felt that was not just something they didn't want to do. But* ***they couldn't****, they couldn't find that out, because there was no way to marry the Children’s Social Care and Adult Social Care databases.*
>
> *Interviewee 19, Policy expert at Consultancy Firm, Nov 2021*

Therefore, developing technological capabilities to overcome the challenges of silos such as the lack of horizontal coordination and a failure to communicate and to share information

---

[80] This consultant worked in a ‘public service mutual’ which as part of the alternative delivery model was promoted through the Department of Digital Culture, Media and Sport (DCMS) to encourage and incentivise public service improvement through solutions that tackle social problems in innovative and commercially savvy ways. Public service mutuals were organisations which had left the public sector (also known as ‘spinning out’) but continued to deliver public services using the experience and insights of employees who had a deep familiarity with the local context. DCMS ran the 'mutual support program' and a fund that local authorities could apply to where they would receive a portion of the funding for local project (75% of funding in the early stages of the project, then 50% and then 25%).

would mean linking data and producing integrated, ‘joined up’ systems. The recurring view was that gathering individual pieces of discrete data held in different contexts, rebundling and recombining them would help you see patterns that can be acted upon to prevent adverse outcomes such as child harm or exploitation later on.

Fixing this ‘silo problem’ was presented as the main reason for the development of a comprehensive ‘data warehouse’ built by combining streams of data held separately by the local council and enforcement, intelligence and vulnerability data held by the local police systems. ‘Project Vision C’ was thus born once the database was in place by creating sophisticated machine learning based predictive risk scores for various issues such as the Child at risk of Criminal Exploitation (CCE), Child at risk of Sexual Exploitation (CSE), and the risk category of NEET that is young people Not in Education, Employment or Training (NEET). Compiling data from across silos in different government organisations about ‘everything that was going on’ with those who had interacted with public services produced an all-encompassing ‘one version of the truth’ about a person that represents a drastic expansion of the overview powers of the state. Under this new paradigm of datafication, information gathering as a function of government is transformed.

> *Actually, it really benefits everybody if you bring all this data together. Because you've got so many different bits of information held in siloed systems. It makes it challenging to try and get that full picture of vulnerability when you are looking at a family. So Project Vision C's problem at that point was to basically try and get as many layers as possible on top of each other, to give us* ***one version of the truth*** *at the local authority level by saying, ‘this is what the police know about this person. This is what the education system knows about this person, this is what Department for Work and Pensions, and this is what local authority knows about this person’. So the project Data warehouse was generated. And then that almost gave the impetus for the predictive modelling. It was kind of the next step because we then knew all the people... [we knew] everything that was going on with the people who we knew about. It was then, how do we identify early the people who are going to become the people that were worried about?*
>
> *Interviewee 8, Programme Support Manager, Early Intervention and Safer Communities, Project Vision C at Local Authority, June 2021*

Data scientists and programme managers attached to this innovative project defended the need for a multi-agency view of the young person even while recognising the different remit of the statutory care and safeguarding data held on someone by the local authority and the enforcement data held by the local police constabulary[81]. The data scientist responsible for setting up *Project Vision C* (who now works for an analytics consultancy) described how previously professionals would be 'going in blind' into a case conference with only a partial view of someone who had triggered concern.

> *What tended to happen was you reach a critical threshold, and they call in a team around the family meeting, or a child or safeguarding conference. And everyone turns up with their laptop and the health worker would give her bit of information, education would give their bit, police would give their bit. Why are we waiting until some sort of high-end statutory intervention, before we share what we all know about this person, that's madness! We should be doing it systematically for everyone. So that we don't reach that point. Because if we'd have all pooled our information, they're going to realise that there are a thousand kids over here that we should be worried about as well… That was the kind of the fundamental point of the whole program.*
>
> *Interviewee 10, Former Data Scientist (now works with an analytics consultancy) Project Vision C, June 2021*

While the details of this case study are expanded upon in the following chapters, the capacity to generate a single, objective, distant and neutral view of the family was seen as delivering objective and direct access to a full picture of vulnerability. In seeking to go beyond silos in government digital architecture through new data technologies, there emerges a longing for a picture that is beyond what is humanly possible to aggregate and 'ontologically superior to what is available to our everyday senses' (McQuillan 2018 :254). Data is able to authoritatively deliver this complete picture. This set of arguments for pooling information makes it seem almost reckless to not systematically share data in order to pre-empt statutory interventions and make early offers of help to marginalised families on the basis of predictive risk scores. It is a discursive regime that makes you accept the premise of combining these datasets while establishing the dangers of siloed IT systems where care professionals go in unprepared for a meeting with an incomplete view of the young person they are talking about.

Full picture of the family that is promised is based on aggregated fragments of data held on the household in various public databases and the risk score overlaid on that by the model. This single point quantitative risk score and the composite record on the family is treated as a proxy or a stand-in for an external, objectively measurable reality. This rhetorical strategy insinuates

---

[81] This project continues to approach linking fragments of data held on different systems as an attractive proposition. However, they now have specific rules about not providing the address of someone held on the council database to the police to issue warrants and to pursue a law enforcement agenda.

at a desire for a complete picture of vulnerability that sits outside human comprehension and beyond old, siloed technological capabilities. Data science and the sophisticated techniques of large scale data linking developed by *Project Vision C* produce a whole perspective that represents an epistemic break from previous administrative procedures; offering a new gold standard of knowledge and 'a higher form of intelligence... that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy' (Boyd and Crawford 2012: 663). The concept of dataism (Van Dijck 2014) has been used to describe this 'widespread belief in the objective quantification' and the potential of data to get at a higher order of reality or generate one, definitive 'version of truth' as the interviewee put it. The advocates for this project have managed to frame it as neutral and necessary while the question of whether these databases (enforcement and council datasets) should be combined remains unasked.

The introduction of new data technologies into welfare bureaucracies is legitimised by recasting administrative IT systems as 'siloed', bureaucratic, outdated and badly maintained modes of recordkeeping. A similar rhetorical trope is mobilised by narratives about the need to 'unlock the hidden value' of public sector information by maximising what you can draw from administrative systems. In fact, the impetus for the *Project Vision C* work began as a radically new way of extracting value from existing data. Interviewees trace the origin story of this project back to 2018 when someone from the police sector took up the senior role of the Troubled Families Coordinator at the local council and realised that local government didn't use data in the same way that he was used to seeing at the police. He felt that local government wasn't as 'capable at bringing together information', wasn't as invested in data driven strategy and analysis and that they weren't using it for anything beyond transactional functions. This person then hired a police intelligence analyst to drastically change the utilisation of data in the sector and the project grew.

Therefore, in this project that used advanced machine learning and data scoring methods, the case for innovation was made by contrasting the local government use of data to an idealised version of policing culture where data was valued and prioritised as a resource and continuously used to improve what the organisation does through rigorous analysis.

> *So what you get in Local Authority generally, is management information form of reporting. So the creation of numbers which get published in a dashboard, perhaps or on some sort of scorecard that says, you know, what's the average time for this? Or*

> *how well are we doing against this particular metric? And the police do that, too. You know, how long does it take us to get things done? Or how many of our crimes do we capture suspect for? We do report all of that management information, what they've also got is a big intelligence team, who analyse data to support frontline policing. So they are looking for, what are our trends, who are the offenders we should be going after?*
>
> *…In a way that council produces management information, but they don't analyse anything. They just go, you know, here's a number. But no one goes, alright, is that a big number or a small number? Is it bigger than it was last year? Is it too big? Is it a problem that we've got? What does it mean, what we actually measuring? What can we do to improve it? You know, all that sort of analysis of anything - that they weren't very good at.*
>
> *Interviewee 10, Former Data Scientist (now works with an analytics consultancy) Project Vision C, June 2021*

By modelling safeguarding after policing systems, the aim seems to be to remove perfunctory uses of data for reporting and move towards comparative analysis of the kinds of families that end up needing the costliest statutory interventions in order to strategise and prioritise support. Instead of producing numbers without any context, the project now probes and provides information in a different format. Instead of saying that someone has gone missing, say 18 times, they report that the child went missing 18 times and on 16 of those occasions, there were numerous risk factors directly related to the CSE index (Child at risk of Sexual Exploitation) as generated by the model.

This project is funded through Troubled Families Programme which is a central government[82] grant provided to local councils for identifying and flagging households with ongoing impediments such as long-term unemployment[83], low school attendance and anti-social behaviour. As a payment by results programme, funding was released when the local authority was able to demonstrate results and show that in the families they were working with, education results and attendance had got slightly better; and they were engaged with their local children's centre. The data performance team in most local authorities traditionally functioned in a very hierarchical environment and had the sole administrative function of collecting up information,

---

[82] The *Ministry* of Housing, *Communities and Local Government*

[83] The Troubled Families Programme is defined as "a cohort of families with problems that are being passed on through the generations; that these families have members who are often vulnerable and in crisis; that they often cause problems to others around them, and that, despite huge efforts put in by so many, they absorb public services without their problems ever being fully dealt with"

collating it and passing it on to central government. It is important to appreciate the manner in which the below interviewee sets out how these routine reporting or ‘evidencing’ tasks were automated as none of this added any value to anyone.

> *There was still an element of evidencing work in order to get money and so that that never really went away. But we got very good at that. We got that down to a bit of an art so that... because I didn't want my team to spend all their time trawling through social care records manually. So let's get all the PBR paying by results as automated as possible. And just not waste any time on that, because that doesn't add any value. It brings in money, right. But families that we're working with, couldn't care less how the council gets funded.*
>
> *My team is interested or was interested in technical problems, they're interested in getting data and ultimately sort of trying to make a difference for improving the efficiency of services. We don't want to spend our time trawling through datasets trying to evidence stuff. That is boring, and it's just not exciting. Whereas I think lots of other local authorities never got to a point where they can stop thinking about that. It just became all consuming. And they'd have two or three people full time, manually checking claims, evidencing stuff to audit teams, and it just never... Yeah, that's not what it's all about.*
>
> *Interviewee 10, Former Data Scientist (now works with an analytics consultancy) Project Vision C, June 2021*

7.3 Conclusion

By identifying the discursive devices used to justify the introduction of new data technologies, this chapter situates its embeddedness in a long-standing culture of measurement, quantification and a rhetoric of public sector modernisation. We have seen how tenuous and novel methods such as user experimentation and live iteration are legitimised by drawing on the discourse of ‘Evidence-based’ policy making and the ‘scientisation’ of public sector functions. This discursive regime, aimed at persuading people that some types of evidence are better than others, displaces the situated judgement and practical wisdom of experienced frontline professionals; while drawing on economistic rationales around efficiency.

The gradual normalisation of datafication in sensitive social policy fields through the ideological legitimation of continuous tracking and various new data streams as good basis for decision making in policy is thus accompanied by assertions that differentiate practices from the private sector. Large scale data linking and decontextualization is justified by rhetorical choices that present them as necessary given the siloed, decrepit state of government IT systems. Nonetheless, rationales for data driven technologies often start with the claim that there is value to be extracted from the ‘raw resource’ of data trapped in these outdated systems.

The case for the introduction of these technologies into government taps into similar narratives by referencing the public sector's information-intensity. It is important to note the unique role of government as the largest 'collector, user, holder, and producer of information about citizens, and organisations' (Heeks 2005). Administrative records are described as 'data holdings' with latent value that is lost when left buried and underutilised within bureaucratic archives. These narratives are used to justify a range of innovations aimed at unlocking data holdings and activating radically different ways of extracting value from data that have significant, ongoing impact on restructuring the logic of social care.

Chapter 8

Reflexive Dataism: Ambivalence and positionality of the ‘data worker’

The aim of this thesis has been to map the extent of datafication in sensitive policy fields such as child safeguarding and welfare provision by focusing on how complex phenomena are reduced to measurable types. Regardless of the variations in their context and background, people are increasingly being identified based on whether they fit into computational categories and data templates. By focusing on two similar institutional contexts at different levels of government, a major contribution of this thesis is the fine-grained descriptions of the models used in these data driven systems that have been hard to access; let alone evaluate up until now. We have seen how new kinds of empiricism has emerged from the acceleration of processes of datafication in child protection where ‘at risk’ models are built out of datasets that combine statutory and policing records. The extensive use of proxies in central government structures the interaction of citizens in tremendously different ways.

The previous chapters have considered the data practices surrounding this profound shift, situated them within their respective organisational contexts and demonstrated the policy rationales that have shaped them. Given the descriptive focus of the research questions, we have seen how the initiators of these projects have an expansive data grabbing vision and employ nuanced justificatory discourses to legitimise their techniques. The thesis clearly shows the advance of datafication with young people and benefit claimants being reduced to scores, and proxies with group characteristics projected onto the individual in ways that can create very real allocative harm. During the course of this in-depth look at the two case studies, what has emerged as ground-breaking and new; that is as yet unexamined in the literature is the manner in which these data scientists see their role in conjunction with these processes of datafication and how as the ‘humans in the loop’ they pause, reflect and adjust their actions.

I find that data professionals or the ‘data workers’ (Kennedy et al. 2015) who work most closely with these technologies display a nuanced understanding of the messy process through which datafication is progressing and interacting with older administrative systems. They don’t have the blind optimism and rash confidence of the tech evangelists who thought of data science as a panacea; but rather feel that the problems they encounter can’t be fully explained or solved by data. They also are keen to position their work as ethical and as self-aware of the limitations of their methods. The overarching finding is that a discourse of *reflexive dataism* coexists with the determination of these data workers to implement and improve their projects and defend them from outsiders.

This chapter presents this key finding through reflections from the day-to-day experiences of data scientists as they acknowledge and grapple with the complexities of their work. Conducting empirical work within a pandemic made it hard to fully access the frontline users of datafied models such as social workers and job centre coaches. This meant that the research question (RQ 4[84]) addressing the role and scope of expertise of the human in the loop could not be investigated as thoroughly as planned. However, data workers and how they see their role in conjunction with processes of datafication emerges very clearly in the case studies. In this chapter, I demonstrate how exploring their reflexive capacity, their ambivalence and struggles with implementing these data systems in a fast-paced regulatory space would push the emerging field of Critical Data Studies (CDS) further. Reflexivity of data workers has been a latent theme so far in the case studies that I want to bring to the surface as I come to the conclusion, and also outline in a limited way the answer to the final research question. I also present my conceptual contribution to the field (with obvious caveats) and will recap the overall findings from the empirical chapters and the main contributions of this thesis in the following chapter. Although the primary findings from the empirical work have already been presented, I will here give some further quotations from my interviewees in order to ground the thesis's conclusions and contributions to literature.

8.1 A ‘bit black boxy’: from the perspective of data workers

A major point of contention within the theoretical debate relates to how datafication frames those implicated in its processes as passive (Pybus et al. 2015). Broad critiques of datafication focus on ‘algorithmic power’ and the thrust of early research on the topic implied that data driven systems are opaque, all-encompassing and universally oppressive. The quantification of the world into data that can then be processed, repurposed, scored and analysed in remarkably new ways led to a focus on the subjects of datafication; and how they are constrained by the

---

[84] ‘As a result of these changes, has the role of the human frontline worker shrunk? Is there now a much diminished scope for their expertise? How do they see their role in conjunction with these processes of datafication?’

biases of these systems (Noble 2018) and subsumed by large scale algorithmic surveillance technologies that have accrued immense power (Ruppert et al. 2015, Zuboff 2019).

This means that in the literature, it is hard to see the people caught up in these systems as anything other than passive generators of vast amounts of data who are acted upon and experience various constraints on their ability to achieve objectives. Kennedy et al. (2015) advocate for the use of the term ‘data workers’ to open up the conceptual space to explore the possibility of agency when dealing with processes of datafication. By thinking about data workers or the people involved in the work of actively ‘making data’, the focus shifts from datafication in terms of power and domination to empirical accounts of how they see their work; and make sense of it so as to act within it. Data scientists employed by these departments can thus be thought of as data workers who make ‘organisational adjustments to accommodate the rise of data’ (Kennedy, Poell and van Dijck 2015: 2) in an active manner. There have been calls to explore the tension between the dominance of structures of datafication, the possibility for agency and the spaces in between. Agency and reflexive data work in this sense would involve exploring how the individual makes sense of their social world through reflection, how they give an account for what they have done and construe their choices and plans in response to their lived experience (Couldry 2013). Despite identifying a key stalemate within the conceptual framing in the critical and structural analysis of datafication, the few empirical accounts of data scientists that have emerged paint them as enacting algorithmic power; and as cogs in the larger machinery that circulates and produces data leading to deeply unequal power relations. There has been very little space to explore how they see their role in conjunction with these processes of datafication which has a large impact on where we go from here.

As such, they are thought of as ‘arch-positivists’ (Tanweer 2018: 102) and data fundamentalists (Crawford 2013) busy reducing the multifarious lives of human beings into stark categories and metrics (Boyd and Crawford 2012, Adams 2016) that are amenable to reductive, quantitative modes of analysis. They are portrayed as blindly assuming the neutrality of their methods and firm believers in data as fixed representations of the world; and as objectively and completely reflecting reality. This would extend to a confidence that the large quantities of structured and unstructured data that is generated about service users and service activity as a matter of routine simply when one accesses help from welfare agencies, mechanically and objectively indicates need and ‘at risk’ behaviour. During the course of this research, detailed reflections from data scientists implementing automation projects show how they don’t see digital trace data as objectively and completely being able to represent human sociality and offline problematic behaviour. They are sceptical about the ability of data from the digital service to capture everything that counts about vulnerable welfare claimants and sometimes even articulate how underlying social problems are rarely fully represented in data.

> *We are not using data in that way***. Because most of these problems are not problems that can be fully explained by data.** *These are problems that go a bit deeper than behaviours. And, behavioural science is incredibly interesting and incredibly difficult. Data science is certainly related to behavioural science in the experimental context but there is way more to it than that.*
>
> *Interviewee 1, Senior Data Scientist, Sept 2020*

Despite spending the first part of the conversation vehemently defending the use of A/B testing and other data intensive methods, the above interviewee clearly understands the limitations of his work. Data only provides a partial explanation and the problems he faces are structural ones that are beyond the behaviours represented by the data he handles. This stands in sharp contrast to the certainty of those who evoke data plenitude and the exhaustibility of datasets[85] to assert that with enough data points one can truly know the world rendering it transparent and manageable. Principal data scientists from the second case study also admit to the fallibility of their modelling ‘as it all comes down to poverty’. Whether by virtue of the records they have on file that are aggregated into their risk models or because of the kinds of measures they are trained to look for, the indicators of vulnerability they use all generate the same list of names where CCE victims look identical to CSE victims.

> *…the fact is, is that all of these vulnerable, vulnerable things look similar. And so your CCE (Child Criminal Exploitation) victims look like CSE (Child Sexual Exploitation) victims look like NEET (Not in Education, Employment or Training) victims that look like people who are in areas of high domestic violence…* **Because it all comes down to poverty.** *And it all comes down to, you know, the common risk factors in the house around, poor school attendance and young people going missing. And the fact is, is that we look at some of our models, if you just did that scoring method, with CCE modelling, the CSE model would look very similar, because crudely all of these young people are experiencing similar vulnerabilities.*

---

[85] Where n=all rather than handling a sample from a population

Interviewee 8, Programme Support Manager, Early Intervention and Safer Communities, Project Vision C at Local Authority, June 2021

This interviewee admits that by just using the index score which is a compilation of the frequency, recency and gravity of flags against young people's names on the statutory dataset, they would get the same list of names in each model. This is why they add on the machine learning model and the network score to narrow things down. He tries to improvise and get around this by treating these models holistically, because risks which might manifest as CSE risk one week, might show up as CCE risk the following week[86]. Therefore, these advanced predictive models are not scientific interventions that are able to identify and prioritise risk in a precise manner. This reflection depicts a clear shift from the earlier breathless optimism surrounding the transformative potential of data that Van Dijck (2014) eloquently captures in the concept of 'dataism'.

We see how contrary to the presumptions of the body of emerging scholarship that has come to be called Critical Data Studies, these systems in the particular context of government departments in the UK, are not necessarily populated by data fundamentalists with naïve and rigid perceptions of the social problems that they are measuring and analysing. They are critical of their own tools and are often acutely aware of the shortcomings of their data. In the empirical chapters focused on justificatory discourses as well as the one that presents findings from the central government case study (Chapter 6), we see how the rationale of efficiency in terms of time savings is often used to justify the use of data science methods. In fact, the emergence of data science overlaps with a historically situated public sector modernisation discourse that is preoccupied with performance management, audit culture, and the evaluation technologies of targets and outcomes. Measuring time spent on a task is therefore crucial to determining usefulness and efficiency given the overarching logic of performance management. We have seen the development of sophisticated digital proxies such as drop off rates and dwell times that are used to check if the digitisation of a particular workflow task has led to tangible savings in time. Even within this context where gathering data and evidence on time spent is essential, the below account from a senior manager shows a critical, circumspect and nuanced

---

[86] "...and it is better to treat it holistically rather than [saying], Okay, we're just going to focus on CCE kids, we are not worried about the CSE kids, it is someone else's job to do that. Because that's when you get to the risk of young people falling from the net." Interviewee 8, June 2021

understanding of how time spent as a metric might not necessarily be measuring what they want to measure.

> *And some of our services have eligibility questions that will come up and say actually, because of this answer, you're not eligible for this benefit. So we would expect to drop out on those sorts of pages. But on the pages in between, we don't want to see that drop off behaviour, because it potentially means that there's a problem... But depending on what the goals are of the service will depend on how we measure whether the service is doing well or not.*
>
> *So, for child maintenance apply, there will be things like completion rate, so the amount of people that are getting through to the end. We probably will look at time… but time is,* ***time is always a bit of a strange metric to use, because you can't control what is going on in the users environment.*** *So it might take 10 minutes to get through the form, because it takes 10 minutes to get through [each section of] the form. Or it might only take five minutes to get through the form. And I spent five minutes dealing with my child who was eating breakfast [with me] so… So you can never tell with time, whether it is the actual time that is taken or not. So, we tend to use that carefully*
>
> *Interviewee 12, Senior Manager – Data Science group, July 2021*

This senior data scientist underscores the need to carefully interpret time spent on submitting a form digitally as it is now possible to gather data on time spent on each task and each screen of the online application with unprecedented precision. Knowing why there are variations in time spent is more difficult; and they still operate based on computational guesses around what correlates to slowness. This might include the inference that the question is worded in a manner that is distressing to the claimant and is causing them to take too long on a particular section. Time metrics are a key measure of whether the service is doing well regardless of what the goals of the service are. However, there is clear acknowledgement that they can’t control the environment of the claimant and what they think might be a problem leading to a longer time spent on the application could just be a mistaken assumption. Data on time is a limited indicator of the actual problem and the interviewee goes on to explain how they also measure the duration of calls to the call centre. They want to reduce the overall number of people who drop out of the online application submission process to then call up the human call centre agent. At the same time, they want the calls that come in to be for people who really need support with unusually complicated questions and the ‘edge case’ circumstances referred to in Chapter 6. They have ways to measure average call duration which they want to see increase as an indication that the easy problems that require short conversations are being solved online and those with complex needs are able to get through and have lengthy, rich conversations with people at the call centre.

Even with the admission that time spent is a crude measure, it is the only routinely collected administrative data used to assess performance of the service. Experience of the claimant in terms of accuracy of the determination by the agent, transparency and feeling of being treated fairly and with due process and dignity would be massively harder to capture in the form of data. The focus on response time and call duration data, even with the reflexive awareness of its limitations, obscures this fundamental fact. Repeated interviews show how despite not falling prey to the pitfalls of naïve forms of dataism, there is a kind of reflexive dataism at work when administrative concepts and social categories are operationalised in data driven infrastructures.

Interviewees from the central government department which has a more formal culture of learning and corporate knowledge sharing mechanisms, talked about intentionally building an innovation lab where they are using synthetic data to learn about their systems. Synthetic datasets are artificially generated dummy datasets which mimic the properties of actual user data but are not real data held on the system from claimants. It was still at very early stages; but they were keen to use synthetic datasets to understand potential biases encoded into the system and run them through it to check if their algorithms handling these datasets would treat certain demographic characteristics fairly. Another team that was in charge of analysing error data discovered that if a claimant had written their name in a field in any application form with a space on the end of it, the system triggers an error as this can’t be passed into their database. An analyst subsequently wrote some ‘clever regex’ to pick out the special character out of the name field without collecting people's names for privacy reasons. The character causing the problem that wasn't an A to Z letter was then automatically trimmed and removing trailing space or a leading space in this manner reduced the amount of errors by 80%. They still had a problem with some characters, which remains a significant problem generally, because ‘people's names are people's names’ and are required for processing benefits and they should be able to accept any sort of character in a proper name. They then shared this advice across the whole service and put it into their design systems which is a library of ways that the website should look and work.

> *So now when services build, they can use that learning from the work that we've already done... But in having that conversation, we also managed to have a conversation about the database, and how the database works and talk about which characters we might*

> *need to potentially strip out of fields, so that they can go into the database; or changes that we might need to make to the database to enable more characters to be put in there.*
>
> *…And we've been able to learn about our database and potentially make even more improvements to our design system so that we deal with more of those problems up front.*
>
> *Interviewee 12, Senior Manager – Data Science group, July 2021*

This account of learning about their database as a starting point for conversations facilitated by data science rather than treating the system as the final arbiter of all decisions surfaces the layers of contributions from multiple teams that built the design system and data infrastructure over time. It shows clear, conscientious acknowledgement that analysing data is a journey not a destination (Neff, Tanweer et al. 2017). The value placed on the conversation about the database; and using synthetic data and other kinds of analysis to talk about and understand how it works demonstrates how the use of data is always mediated by interpretation and stories that allow it to have meaning across teams with different expertise and practices. The application that captured the name field was probably designed by a different team with different design principles and values that are not always fully documented and passed on. Data scientists see these datasets as 'sites for conversations' (Fiore-Silfvast and Neff 2013) that lead to new questions, solicit more input, and make transparent the context and assumptions that go into the initial design choices. Diametrically opposed to the notion of being scheming 'data fundamentalists' that see large datasets and predictive models as 'reflecting objective truth' in a manner that is beyond the pale of human comprehension and what is humanly possible to aggregate, it emerges that they in fact see large datasets as starting points that support decision making and not a finished, insular system (Neff et al. 2017, Seaver 2017).

It is worth pointing out that in thinking about synthetic datasets as being able to expose systemic bias and using regex to remove all personal data, these respondents still display a considerable faith in technical fixes. Bias which can be eliminated by a 'scrubbing to neutral' approach betrays a lack of structural understanding of these systems. Nonetheless, we see how data is tactically used and shaped by local practices. The manner in which these data scientists are conducting open conversations with different stakeholders, reflecting collectively and actively figuring out how to have those reflections and conversations reproduced in other team settings shows them wrestling with multiple priorities and divergent interests. Data work involves social and organisational work and a series of negotiations.

Investigating these empirical contexts have shown us a much more reflexive, circumspect and conscientious practice of data science than is portrayed in the critical literature.

> *…Pick me up people who look similar…. But* ***it's a bit black boxy*** *in that… then we'll work with the other two methods in this system to produce a score at the end. And what we found as a kind of reflective point of view to include is, it's not good to have that. You want... explainability needs to be the utmost because if someone says to you why was their score this, and you say the computer just works it out, then that is when you lose trust.*
>
> *Interviewee 8, Programme Support Manager, Early Intervention and Safer Communities, Project Vision C at Local Authority, June 2021*

If critiques of the big data turn have problematised rampant positivism where data is taken as a given, another key insight was that data scientists inhabit a technocratic world that remains ‘black boxed’ and inaccessible to public scrutiny or wider democratic input (Pasquale 2015). The manner in which the above interviewee talks about his child safeguarding risk models[87] as being a ‘bit black boxy’ incorporates the language of the critique, and echoes the theoretical terms with a kind of resigned irony in an attempt to parody the enthusiastic and simplistic belief in data driven modelling.

This comment was made at the end of an interview without me prompting him or introducing the concerns with his profession or Pasquale’s characterisation of the opacity of proprietary models into the conversation. After three years into implementing datafied risk measures and mainstreaming them into the older administrative workflow, being unable to explain the single point numerical risk score aggregated from the three models is what he identifies as creating mistrust in the system. Data scientists and practitioners are increasingly sensitive to public perceptions of data science and algorithmic systems and are able to use the lexicon of the critical literature to reflect on their practice.

8.2 ‘Not as Foucaultian as people think it is’: ambivalence and ethical posturing

There is an unmistakable sense of ambivalence when interviewees reflect on their work in relation to existing concerns about algorithmic control and the ethics of big data. Most of the

---

[87] The second model scores what ‘looks like abuse’ using historical datasets

data scientists I encountered are idealists who see their work as doing good and are generally well meaning in the sense that they have given up well paying private sector jobs for a key piece of public service work. Majority of data collection was conducted during the second and third covid lockdowns in the autumn of 2020 and the first half of 2021. In the lead up to the summer of 2020, schools were shut down with isolation mandates and had to shift to a remote, online mode of delivery. In person A-level exams that determined entrance to universities had to be cancelled owing to emergency measures and following the use of algorithms to round off end of year results there were unprecedented, widespread protests calling for the Department of Education to 'ditch the algorithm' (Amoore 2020)[88].

Media coverage and persistent Guardian exposés over the last 5 years of controversial Early Help Predictive modelling in local government led to ongoing, lively public debate about black boxed algorithmic systems. This was compounded by increasing concern about contact tracing apps, publicity around 'vaccine passports' and the use of the NHS app which had updated data on covid test results, date of last infection and vaccination status as an identity document to access venues. The heightened awareness caused by these visible protests meant that data workers were forced to confront and work within the context of rampant worry about algorithms and the ethics of big data. So much so that initial conversations negotiating access and early parts of interviews contained constant reassurances that their work was ethical. I noticed that before rapport had been established, respondents carefully considered their words and consciously avoided the use of the word 'algorithm' or 'predictive modelling' as they were now loaded, controversial terms. Algorithm was replaced with 'decision support system' or just the word 'data' that retained its interpretative flexibility; and constant reiterations that emphasised that there would always be human involvement in their projects.

> *It is just… it is a complex system, I don't think… I can't imagine any process that is entirely governed by a computer. In any part of government, it is all based on policy and legal requirements. It is not just a computer making up and saying (unclear) what to do****. So, I just don't think it is as Foucaultian as people think it is.***
>
> *Interviewee 1, Senior Data Scientist, Sept 2020*

---

[88] Chants that used expletives and placards that read 'F*** the algorithm' were prominent – as were banners that read 'The algorithm stole my future'.

Data scientists see their work as rendering legal and policy requirements into code. They believe that even while digitising workflows and changing the nature of welfare and social care support, there is sufficient oversight to prevent them from creating automated processes that could harm vulnerable benefit claimants. I heard that drawing on the expertise of policy colleagues from a different part of the organisation and those with long years of experience helped them think about how things were done in the pre-datafication era and insulated their work from obvious mistakes.

User researchers are incorporated into the work of data science teams although these researchers do not use representative samples and often rely on market research panels for sampling. They do not produce foundational or non-applied research that generates new hypotheses and problem definitions about areas that need attention; and only run projects to check if the new data-based modification is working properly for their target population. It is unclear whether there are mechanisms in place to incorporate the post-hoc feedback about how some claimants' needs are being systematically excluded by datafied technologies and how much of the user research can be used to produce major changes. There is an intentional effort to include multidisciplinary teams in major service redesign projects so that different perspectives other than what is the most convenient course of action from a technical point of view is heard[89]. The above interviewee shows how they are aware of the dystopian view that similar to Foucault's theorisation of the use of technologies to discipline, punish people and reproduce capillaries of power, the proliferation of automated data-based processes signifies a worrying shift. He tries to be reassuring and defend his design choices by saying that no process in any part of government will ever be entirely governed by a computer. As I heard repeatedly when interviewees were pressed to reflect on the fundamental changes activated when datafied models are used to identify and verify claimants and segment populations based on risk, he also reiterates that there will always be a human in the loop.

Popular imaginations about data workers or the 'humans in the loop' of data driven systems revolve around their ability to override automated decision making and with 'kill switches' that can disable automated features. My research has shown how political imperatives have led to the architecture of the system as it stands and how we need to think about data infrastructure as an interconnected system and as situated within organizational contexts; rather than about

---

[89] This also means including diverse team members in the department with lived experience of disability or disadvantage during the Discovery and Beta phase of the design process as a way of sampling using the human resources in the room.

algorithms as technical procedures that produce individual outputs to which someone can say ‘no’ and thereby override them. The earlier account of negotiations and cross team communication (Interview 12) demonstrates how very interrelated these roles are rendering the question of whether one person being able to individually override a single automated decision recommended by the system means anything at all. However, interviewees are aware of this moral panic around automation and constantly reiterated that there will always be a human in the process and it will never just be the computer making decisions.

Needing to demonstrate that they are acting ethically, and being cognisant of societal perceptions of data driven technologies while having to perform their roles within departments that have a top-down mandate to implement an ‘automation by default’ vision proves to be a tough tightrope act. Fear of algorithms and data science closes up spaces for discussing these changes that are inevitably going to happen while at the same time causing anxiety, deep ambivalence and a kind of ethical posturing amongst data workers. This is best exemplified in the quote below which followed an answer where he apologised for not being able to answer my question in a full and frank manner.

> *Interviewee: Even that I think would be difficult. Sorry, yeah, unfortunately it is an interesting period for data science. Because everyone is scared of data science.*
>
> *So, we have to kind of predict, kind of pre-empt all the concerns before they even happen and try and stop people from imagining these nightmare scenarios. Do you know what I mean? Unfortunately, it creates a kind of incredibly… Everyone is like super cautious and a lot of problems that data science might solve to do identity or fraud and things like that is something that… You don’t want people to know how we solve them. Because that would be useful for them [to cheat and defraud the system].*
>
> *Interviewee 1, Senior Data Scientist, Sept 2020*

The most detailed reflexive conception of the role of a data worker came from the principal data scientist who designed the scoring models in the local government case study. He had an impressive grasp of the key concerns being raised by critical scholars, followed their research, and spoke about trying to ‘proof’ his work as the ethical debates around predictive analytics and data science in government heats up.

> *I also wanted to give us a bit of insulation, because I was always thinking, there's a debate heating up around... So, I read the Kate Crawfords, and the [unclear], the Trevor Paglens of the world on Twitter, I try and keep track of their work. I don't have as much time to read papers as I would like. But I want to know the headlines, the kinds of things they're talking about, so I can try and find new things in my work. So I always knew that that ethical, legal debate is heating up in the background.*

> *So, I wanted to protect us against the world where maybe at some point, the decision to turn around as the public sector is not going to be using ML tools. If it was just index score, it doesn't really fall into the second set of arguments, it still means that we can produce a prioritized cohort. So by having three methods, I thought I could protect us a little bit against that if it ever happens because we still need to prioritize who we are working with.*
>
> *Interviewee 14, Business Intelligence Developer, Project Vision C, Aug 2021*

He sees the index and network score that literally adds flags against a young person on statutory databases to denote risk as a straightforward use of data and felt that without the predictive ML aspect of their score, their work met the highest ethical standards. By recalibrating his work based on the debate around data ethics voiced by Kate Crawford and others who used Twitter and such public, non-academic fora to broadcast their critiques, he felt confident that these models would withstand scrutiny. He also followed the ethical concerns as it spilled over to the legal domain with debates around procedural fairness in administrative law by keeping abreast of ‘headlines’ even if he didn’t have time to read the full papers. He felt that if at some point following public concern as the ‘debate heats up’, if there was a concerted decision that the public sector would no longer use machine learning tools, the resources poured into developing the datafied risk modelling would not be wasted. Even while trying to pre-empt criticism, the scoring model is framed as being able to withstand accusations about overreach in terms of data collection, processing and predictive labelling that potentially circumscribes the life chances of young people deemed to be ‘at risk’ (rather than gets them support).

Working within a fast-paced regulatory space, I heard from others about the stresses of having to balance the needs of changing legal and data protection requirements with public concern and their ongoing work. Data Scientists embedded in poorly resourced public sector departments had to produce organisational policies and position documents in response to national standards for data use that were introduced with the adoption of GDPR, the National Data Strategy (DCMS 2020) and other centralised legal instruments. There was a sense of exasperation where another data scientist attached to the same department talked about how there weren’t ‘any clear black and white legal legislation’ that sets out what you can and can’t do with data.

It is always open to interpretation and was a grey area where people are always going to push it too far until there is greater clarity in the regulatory and legal guidelines. As a team, they presented their innovative data driven approach over 300 times (since 2018) to other local authorities and have a practice of presenting to academics and civil society organisations to understand societal concerns with their work. In the absence of such clarity in data protection law, it came down to individuals to be open about what they were doing, solicit input and concerns in a tentative way and have organisational ethical frameworks that locally interpret national laws.

> ***Interviewee:*** *So we're really open about what we're doing, and engage with anybody that will listen, for better or worse…I think if you have an honest conversation with people about this is what we're doing. This is how it is been done. This is why... then most people don't have an issue with it. In fact, most people think you're doing it anyway. So all the kind of disclosure stuff we've done, we've always thought was obvious.*
>
> *Of course [emphatically] my social worker should know when I turned up... when my kid wasn't at school. The fact they don't is mad… And there are opt out clauses you can put in... the whole time I was there [at the local government team as he has since moved to a new role] I don't think we had a single Subject Access Request through GDPR. We didn't have a single 'right to deletion' or nobody asked to be removed. And of any local authority or police force we had far more publicity about what we were doing than anybody else.*
>
> *So I think it's all a bit of a grey area legally, with any GDPR data protection. And especially the new EU legislation is absolutely rubbish; it is not clear at all.*
>
> ***AT****: it is based on general principles, yeah, yeah.*
>
> ***Interviewee:*** *…because there isn't any clear black and white legal legislation, the answer comes down to you as an individual to have a principled approach and to say, this is what we're doing and why…* ***And until there's a bit of legislation that says, this is actually what you can and can't do, you're always going to get people that either push that too far, or aren't happy with it.***
>
> *…****I mean there is ethics frameworks for days, like everyone's got one.*** *Yeah, you can tick which ever one you want. You want a bit of, I think, in order to silence or have these conversations you need someone like, well probably the EU, because we've left. We need a bit of legislation that clearly says in black and white, what you can and can't do. We used to have so many discussions with people about the Digital Economy Act, our interpretation of legal gateways and how to make databases and stuff - it is just too grey. There's one person who reads it one way and one person reads it another, you just can't agree. Yeah.*
>
> *Interviewee 10, Former Data Scientist (now works with an analytics consultancy) Project Vision C, June 2021*

This decision to demonstrate how they are locally interpreting data protection guidelines is a best practice example of transparency developed by this local government team. Ethical frameworks remain voluntary. Even in areas where there is data protection legislation, local realities of data use, curation and recombination that defy contextual integrity (Nissenbaum 2009) mean that these are open to interpretation and depend on the commitment of the organisation to be reflexive, open and solicit contrary opinions. The pressure to function within the legally grey area where there isn't sufficient clarity means that they expect the young people who are being categorised based on risk scores to have sufficient data literacy to take the initiative to opt out of these systems. This is a rather impractical expectation as the subjects of these models don't know enough about how their data is being captured, repurposed, and aggregated to file 'right to deletion' requests with a kind of cool-headed rationality and entitled confidence.

We have seen how rather than retaining a naïve faith in data technology and rigid limiting assumptions about vulnerable welfare claimants, complex need and human life, data workers are driven by mundane project realities and heightened awareness of the complexities of their role. These final sections have considered their ambivalent positionality as humans in the loop both implementing these systems, understanding their limitations, and defending their design choices. When asked to directly respond to concerns about the growing use of data driven technologies and automation processes in sensitive policy fields in terms of their disadvantages, the musings of the following senior manager are instructive.

> **Interviewee:** *[laughs] There are implications in terms of staff numbers, does Department A need to be as big as it is? …but that doesn't really fly with me. Personally, I think if I made myself redundant, by doing all the IT stuff in the world, and then had to go and teach people violin, that is great! So that is a different problem, that is a post capitalist society. If we make all the decisions by computers, rather than allowing some decisions made by people, it will be... cheating the system will be different.*
>
> **AT:** *Did you say difficult?*
>
> **Interviewee:** *Different. So you sometimes read sensationalist newspaper writing saying, because this decision was moved to the computer, either there's been this level of fraud, or these people who deserve don't get the benefits. To me the computers should be good enough if that is the case. It's not that those things happen, that the people who deserved it didn't get it. These days, the fraud, and the error is different. So those two potential issues. Personally, I don't buy into either of them: reducing the size of Department A; making decisions by computers changes a threat, it doesn't add threats.*

Interviewee 2, Senior Delivery Manager, October 2020

After a long pause following the question about possible disadvantages, the interviewee looked slightly uncomfortable and painted a utopian picture where after full data based personalisation of services in a post capitalist society, he might possibly make himself redundant. The main disadvantage he identified was that increasing automation made human roles redundant in a public sector department as big as his. When addressing the moral panic and heightened concerns about data driven technologies, he characterises them as sensationalist media portrayals and emphasises that the problem is not with the technology; and that automated decision making changes a threat and doesn't add it. Rather than being pessimistic over alarmist claims that algorithms were taking over, he felt that the complete implementation of his vision might lead to him having to teach people violin and work as a musician sometime in the distant future as the computer takes over all mundane and repetitive tasks.

In concluding this thesis, I have chosen to foreground the reflections and subjectivities of data workers rather than hold them responsible for the systems that they help produce and maintain as if they were all powerful and singularly responsible for the overreaches of datafication. Owing to this analytical choice, we are able to see from their perspective how the mundane realities of bureaucracy such as project delivery problems shape the messy manner in which datafication is unfolding. Even while documenting their reflexivity throughout this chapter, I do not seek to imply that they are reflexive enough to shift the dialogue in the whole agency but only explore how they differ from the blind optimism and rash confidence of tech evangelists or data fundamentalists as portrayed in the literature.

Critical Data Studies has articulated the oppressive features of algorithmic power so well that there have been very few attempts to develop capacities to understand and intervene around processes of datafication by thinking about data in relation to agency[90]. In focusing on the day-

---

[90] There have been some early attempts to do this in work looking at the role of data activists (Baack 2015) and the possibilities of 'algorithmic antagonisms' to disrupt and subvert power (Heemsbergen, Treré and Pereira 2022). There is also a large ESRC project titled 'Living with Data' (Kennedy, Taylor, Oman, Bates, Medina-Perea, Ditchfield and Pinney 2021) looking at how ordinary citizens respond to the realities of living with heightened processes of datafication. At the time of writing of this thesis they had only released early survey findings showing overall high public concern with data and results from a question that hypothetically asked all respondents (not just welfare claimants) if they were comfortable with the kind of data driven identity verification technologies used in relation to Universal Credit. Most respondents were largely comfortable with identity verification technology but were less comfortable with the possibility of government departments checking whether people logging into accounts swipe their phones in the same pattern as they normally swipe it as a future extra security

to-day experiences of data workers as they acknowledge and grapple with the complexities of their work, we are able to see the variable, active and contingent ways in which they exercise their agency and reflexivity. We see how data driven systems are not necessarily populated by data fundamentalists, but rather by diverse and ambivalent characters, who work in contexts that are not necessarily stable or coherent but with ‘meanings that cannot be simply read or guessed at, but which must be found empirically’ (Seaver 2017).

This shift in focus is important as developing capacities to understand, unpack and intervene into datafied black boxed systems is more important now than ever. The existing critical literature has articulated the oppressive features of algorithmic power at the expense of investigating alternative epistemologies of data that engage with the possibilities of reflexive action in the age of datafication. While still premature to sketch out definitively, future research developing this analytical capacity could involve working with the tools of datafication to raise questions about the aims of these automation projects, reclaiming data literacy as a means to open them up, highlighting the social problems that remain outside the scope of technical solutions, inviting democratic input and at some future point recalibrating these systems based on such input.

If these ideals of recalibrating the automated welfare system in a manner that includes the most extreme ‘edge cases’, provides dignity to disenfranchised welfare claimants and incorporates and values democratic input above technocratic and political rationales, are to become a reality, the views of those who are closest to these processes are vital to document. Mapping the subjectivities of those who are in charge of implementing automation projects; and documenting their reflexive capacity in terms of how they see themselves and their work are key to understanding if they are able to make fundamental changes. This is especially true as the terms of the ethical debate are framed such that widespread aversion to algorithms and automation is closing up spaces for dialogue; and the opportunity for the voices of critique to be used to improve and transform data science is slowly disappearing.

---

check. People in receipt of Universal Credit were more comfortable with the checks described than people not in receipt of Universal Credit.

Chapter 9

Conclusion

Summary of Theoretical and Analytical contributions - A differentiated understanding of Data

The main way in which this thesis contributes to the theoretical conversation is in how it develops a differentiated understanding of data and maps the incursion of processes of datafication into social care and welfare provision in a situated manner – valences and meanings of data are contextually defined and have to be empirically found. The empirical exploration of the datafied risk model at the local government level in chapters 3 and 4 showed how data workers adopted a 'data led' approach that computationally queried a 'haystack' or a data warehouse to look for a 'needle' based on measurable types of 'at risk' that drew on what victims had historically done.

The meaning of 'data led' in this case study is specific to this project where they are trawling through a data warehouse that holds full records on what previous victims of exploitation looked like before they were referred to professionals, without looking for anything specific but using the properties of these data signatures to tell them what to look out for. In other words, data led here means going beyond what the social workers know to expecting a picture of the 'needle' to emerge inductively (without looking for anything specific) and for the measurable type for a *typical* child at risk of future abuse to emerge. Using the case of health tracking apps, Fiore-Silfvast and Neff (2013) explain how different data valences emerge as data accrues meaning within a certain organisational and institutional context. This could mean data as 'truthiness' which makes an illness real, visible and more true than a vague feeling; or data as actionable as leading to immediate insights; or data as producing self-evidence where you can diagnose yourself using the logged data on vital stats, heart rate, weight and so on and draw causal links without a doctor's expertise (Fiore-Silfvast and Neff 2013). In my case study, data as generating what 'looks like an abuse victim' and producing risk scores based on similarity to that prototype, is locally defined and produced through negotiations, organisational and social work. It incorporates the assumption that previous victims' circumstances will be repeated in the case of future child sexual abuse or criminal abuse cases; and that the answer to the problem of early signs of exploitation going undetected by lead professionals is in the digital traces of contact with statutory providers as found in the comprehensive system somewhere.

The second case study (Chapter 5 to 6) is of a department with purview over welfare, pensions and child maintenance policy that is digitising various benefit lines; and wants to increase the capture of data from the service to learn more about the user and support the development of analytical models that are embedded into operational systems in order to capture 'customer insights'. Datafication processes are specific here and institutionally shaped by contextual realities. The valence of data as developed in this central government context relates to producing a composite picture of the real human behind a form or a case by collating data from different departments and personalising based on life-events. Uses and practices of data here are uniquely shaped by the political imperative to cut welfare budgets and a top down push to make them a 'digital exemplar'.

Datafication does not proceed in every context by producing inscrutable insular models, displacing human discretion, bulldozing and replacing earlier modes of working, generating unequal power relationships and reducing all contextual variation and differences in human lived experience into a stark, single point metric. Datafication never progresses the same way in every context and almost certainly does not occur as a smooth transition everywhere. It progresses in a messy manner with the imperatives of datafication interacting with organisational practices on the ground in an uneven way; sometimes changing local routines and sometimes experiencing resistance when frontline practitioners come to terms with the limitations of these new technologies and use buffering strategies to minimise the impact of datafication on their work. We have already seen the existence of a kind of 'reflexive dataism' where there are exhortations to dig behind the data, acknowledgement that outputs of data driven models are sometimes wrong and tactical uses of data where data workers keep abreast of the wider legal landscape to tweak and modify their projects. We have also seen the use of open conversations to understand, recalibrate and adapt outputs from a model; and instances where they have to regularly confront the limitations of data based measurement devices (like time metrics) and discuss and unpack the design assumptions of datasets in order to make it work for them. The meaning, practices and nature of datafication can't be guessed at and has to be empirically found in each context; and by developing these in-depth case studies, this thesis is one of the first detailed investigations of the practices and discourses of datafication in welfare bureaucracies and within the social service delivery context.

Prominent theorisations of datafication in the literature such as in the body of scholarship around surveillance capitalism (Zuboff 2015, Zuboff 2019) or the data colonialism thesis (Couldry and Mejias 2019, Couldry and Mejias 2019) see data as necessarily oppressive; focusing on the extractive process that turns them into economic value while unleashing inescapable surveillance systems. Approaches such as mine situate data in local practices to produce a contextually embodied and differentiated understanding of data. Such approaches are significantly less alarmist and are a much needed empirical check against the dominant rhetoric of algorithmic power (Christin 2017). Rather than focus on data as an abstract notion or the data driven technologies themselves by explicating their technical features and algorithmic procedures, I focus on how they are used - the situated manner in which people choose to use the datafied models, activate interpretations, conversations and discourses around their outputs and draw them into their sociocultural lexicon. The following recap of each of the empirical chapters demonstrates how this approach has produced findings that contribute to the field in distinctive ways. The following chapters 3 to 6 focus on the practices of datafication and develop in depth case studies in a fine-grained manner and is followed by a chapter that delineates the discourses used to justify the introduction of new data driven technologies.

Chapters 3 and 4 report on findings from the local government case study of the datafied 'at risk' scoring model in child protection. This datafied model has been placed at the heart of a bureaucratic workflow in order to prioritise those at high risk using a combined score drawn from an index that counts the number of times there have been flags against a young person's name in statutory databases, a network score that quantifies peer association risk and a predictive score created through a machine learning model. The modelling used in each of these three scores are discussed at length in terms of the classificatory logic it sets in motion, the appropriateness of the datasets used and how in pre-emptively trying to identify harm through data they solidify and reinscribe historical interactions with welfare agencies and safeguarding teams.

Sorting these scores into categories (tier 1 to 3 where tier 1 is high risk and tier 3 is low risk) creates a priority list of names that is automated to update every week based on incoming incidents and appear on the dashboards of children's social care professionals at the local council. This way of computing risk and mainstreaming the datafied model into work routines, has since 2018 unearthed names of young people who were not previously known to council workers as part of child protection plans or existing case work. Owing to the flattening effects of categories, when seeing ‘at risk’ as a compilation of data signals and classifying a pupil based on algorithmic fit would mean that someone’s circumstances which might be no more than a tier 3 low risk appears to be more dire than it is. The predictive score is built through a decision tree ML model based on how similar you are to the undesirable behavioural traits of previous victims of child abuse and data patterns of what ‘people like you’ have done in the past. This predictive model is computed through machine learning process trained on about a hundred current victims whose abuse allegations have been substantiated and work with case workers from Barnados’ (“*We ask what did that person look like when they were 12 rather than 16 [when they were identified as victims]?* - Interviewee 10). Inferring group-based information or population level risk scores and judgements about this base cohort and translating them into absolute predictions at the individual level sets in motion a major classificatory problem. This group-to-individual problem in the scoring of a child at risk of sexual exploitation (CSE) causing a breach of a young person’s informational privacy and informational identity is discussed using insights from Information Ethics scholars.

By analytically focusing on the socially constructed nature of the dataset created when records from the justice and policing systems are combined with statutory social datasets held by the council, I show how socioeconomic disadvantage is reified into digital proxies. In trying to let ‘data lead’ and creating prototypes for hidden risk in order to pre-emptively identify where the future child exploitation caseload is going to come from, the system as a whole can become intrusive when collating sensitive mental health and domestic violence records and can reinforce negative gendered and racialised judgements. I identify how datafication in this case study corresponds to a logic of ‘risk based prioritisation’ where protective factors that are sociological in origin and can’t be captured into a data template are not identified. This can lead to a young person being labelled as ‘risky’ making it hard for a lead professional to adopt a strengths-based approach to providing restorative, pastoral support in a trauma informed way to a vulnerable person[91]. The network score shows how dependent on artificial informational relationships this datafied model is leaving it to the practitioners to retrieve what the score erases and everything that cannot be quantified by the risk-measurement model.

---

[91] Focussing on risk factors through a deficit model and disregarding someone’s strengths and what a young person can do at the start of social worker relationship with a family (whether it be for serious assessments of child maltreatment or in order to initiate a conversation using the Signs of Safety toolkit) leads to a different kind of rapport.

Given the clear class biases baked into these datasets, I elaborate on the blind spots and unintended consequences of this kind of modelling where automated risk categories are being built on systems of exclusion already in place thereby exacerbating them. While studies of predictive modelling in child safeguarding have been much researched in the datafication literature, this is a unique case study that makes significant contributions to knowledge as the existing empirical studies have been of routine child protection functions where a datafied model is used to predict if, say, a child at the age of 2 is at risk of harm at 4 (Gillingham 2015, Keddell 2015, Cuccaro-Alamin et al. 2017, Gillingham 2019). The most famous child protection study is of the Allegheny County algorithm from the United States (Keddell 2019, Vaithianathan et al. 2019, Veale 2019). The cancelled predictive child maltreatment model in Illinois (USA) and New Zealand that have both been much researched traced their poor predictive capability to poor data records on service users and service activity (Gillingham 2016). I present findings from the UK welfare system which is significantly different to the US child protection system which rather than trying to understand family problems, has traditionally been more punitive and is organised around the logic of seeking to identify crime. Furthermore, rather than regular statutory functions such as predicting who would need Child In Need plans or Child Protection assessments, Project Vision C deals with a suite of risk models focussed on predicting Child at risk of Criminal Exploitation (CCE), Child at risk of Sexual Exploitation (CSE), and the risk category of NEET that is young people Not in Education, Employment or Training (NEET). The combination of statutory and enforcement datasets raises new empirical questions that haven't been addressed in the literature so far.

In this case study (and throughout this thesis), I draw on Critical Data Studies' foundational contribution of problematising 'data' as objective, neutral or disembedded from the calculative practices and logics of equivalence that translate the phenomena it is meant to represent into observable reality. I historicise the datasets fed into this suite of risk models by focusing on how the architecture of statutory electronic case management systems come from the Poor Law system from the 18th century and overwhelmingly deals with low-income families, ethnic minorities, single mothers and others on highly stigmatised forms of assistance. Policies such as Troubled Families programme significantly determine what is held on record by councils and mean that administrative datasets are incomplete in a non-random way. In seeking to define the measurable type of what 'looks like a victim' based on data on welfare history and on digital transactions that troubled families have had with schools, courts, and unemployment services, the system imports biases and amplifies in a precise way the vulnerabilities of young people on the fault lines of society. The analysis in the chapters 4 and 5 thus makes a distinctive contribution to the existing studies of predictive modelling in child protection as it makes visible the embeddedness of datasets in the historical and political imperatives that shape their current architecture.

Chapters 5 and 6 adopt a lateral approach and instead of a single datafied model like in the last few chapters, I study in careful detail a series of analytics processes instituted across the department where there is an ongoing push to embed analytical models into operational systems with data being continuously captured from the service to generate granular, real-time insights about patterns and risks. The vision for datafication in this department revolved around using, matching, and combining data points held by various departments to generate personalised prompts based on life events and using data to build an understanding of the real person behind the paper form or case.

By tracing policy choices over the last few decades, I identify a heteronormative data infrastructure that decentres women and single parents. I then demonstrate the adverse effects of fully realising this vision of forming a datafied view of person and a composite picture of their household composition based on data that is not voluntarily provided. Linking a female carer’s national insurance number (NINO) and bank details to her male partner and compiling intimate details such as their childcare custody arrangements, illness, and financial information does not merely simplify citizen interactions but pre-emptively makes decisions before the welfare recipient engages with the department and erases those who don’t conform to heteronormative stereotypes. The category of the household became important during the course of this research especially as it became the unit for implementing COVID measures of isolation, and what the government used to plan the location of testing facilities and so on. The inferences about individuals and their households based on data on the system can now be used for planning purposes and this vision of transforming the way in which people interact with the digital service has clear blind spots. This analysis clearly shows how there needs to be guidelines limiting the secondary uses of these data points.

I demonstrate how optimising the system in this way increasingly targets and redlines marginalised groups without intent. By drawing on the example of the state pension provision which is one of their success stories, I show how continuing financial and digital exclusion issues mean that even the radical disintermediation of the pension delivery system is optimised for those who are digitally savvy, have good credit checks, are debt free and in good standing with their bank, earned a good income over the life course, and have their names on key home ownership documents. In this manner I show how the humane, benevolent anticipatory governance imagined through datafication though signifying a key ideological shift towards a more caring welfare system has many ‘edge cases’ that need attending to. Given the political and ideological threads that created this data infrastructure, it is worth asking if this vision for a more generous, less punitive welfare system can be delivered through data.

These chapters show how digital proxies based on online user journeys generate a ‘calculated public’, a datafied view of different kinds of users and allocates vulnerable claimants accessing the website during a hard time in their lives into algorithmic categories that defy human comprehension. A key contribution of this chapter is the manner in which it links proxies through which one is involuntarily assembled into a group based on clicking behaviour or web traffic to emerging discussions of how there are no legal protections for those excluded by data-based categorisation unless they can be mapped onto an existing protected group. Such linkage is rare in practice (Wachter 2022).

The empirical chapter (**Chapter 7**) on Justificatory Discourses accompanies this in-depth dive into the practices of datafication by tracing the discourses used to justify its introduction. I identify the argumentative threads used to frame complex policy problems as discrete, purely technical problems that can be solved through data science. Policy and governance problems in this sector if opened up for deliberation, democratic input and debate would reveal contentious goals with different stakeholders having different interests and needs. By aligning novel, uncertain methods such as automated A/B testing with a culture of expertise centred on evidence, data science is legitimised. Moreover, new kinds of evidence for policy are rendered palatable, appealing and desirable. I trace the emergence of the Evidence Based Policymaking movement as a historically situated public sector modernisation discourse that views long-standing policy problems as discreet, measurable puzzles that can be solved using technical expertise and more direct and instrumental use of evidence.

Even as the rationalities of New Public Management meant an obsession with performance management, evaluation, and the technologies of indicators and targets, datafication has created new measurement techniques that make a mode of continuous evaluation possible - where it is now possible to gather responses to changes made in real time, know demonstrable impact and ‘what works’ almost immediately. I call this a ‘new kind of empiricism’ that involves constantly checking if something is doing what you want through iteration, the tabulation of real time information on how users are responding to a new digital feature and constantly checking how design choices are faring. Often this involves enrolling vulnerable users like welfare claimants into live trials such as large scale A/B tests without their knowledge. This chapter also includes a brief discussion of the ethical implications of this ethos of experimentation where live users are tagged and routed through different control and stimuli variations to measure differences in outcome under the guise of producing robust and good evidence. These sophisticated data technologies propagate a disproportionate emphasis on measurement and produce underlying changes in the rationality of welfare bureaucracies.

Very little is currently known about the inner workings of such departments in terms of how they allocate resources, produce insights from their digital systems and make decisions. Therefore, a key contribution of this thesis are these empirical descriptions of the use of A/B tests, live iterations and how the digital service is optimised based on large scale data from user experimentation as most platforms are released into the live environment as a minimum viable product with much functionality still to be added. I situate these new data technologies within a long-standing culture of measurement, quantification and a rhetoric of public sector modernisation. The impetus to artificially create these A/B tests and large scale trials that partition tens of thousands of claimants into control and treatment groups where different versions of the digital service feature are shown to different people comes from a need to appear more scientific and demonstrate results and outcomes. It strains the administrative law principle of needing to treat everyone equally when implementing government policy. In documenting the gradual normalisation of these techniques of datafication, I identify a brand of datafied instrumentalism that elevates quantification, precision and measurability over equitability and other principles of good governance.

The analytical strategy of this thesis has been to approach the research problematic in three parts. In each of the chapters, I empirically trace the practices of datafication, then proceed to a brief discussion unpacking the consequences of that datafied architecture in terms of who it is optimised for and who is implicitly excluded. Then, the most significant theoretical intervention of this thesis is in the manner in which I situate these data practices within the social and political pressures around them. In thus heeding Christin’s (2017, 2020) call to stop approaching data technologies in a decontextualised manner I am able to create the conceptual space to trace the often contradictory manner in which data scientists and frontline practitioners talk about the practical realities of their work in relation to the original vision. Organisational culture becomes important and we see a series of negotiations by data scientists who rather than being passive tools of reproduction of algorithmic power wrestle with multiple priorities and divergent interests. By linking to historical measurement cultures and the political rhetoric of cutting welfare budgets and nudging deviant communities, it becomes clearer why these technologies target in the manner that they do. The continuities with previous public administration rationalities become evident as does the distinct fashion in which data is reconstituting welfare provision in a way that has significant implications going forward.

The key point of departure for this thesis is the manner in which it brings Social Policy literature and puts it in conversation with the emerging canon of Critical Data Studies. I am therefore able to ask questions about the origins of the datasets in the child protection risk scoring model and see the emergence of risk based prioritisation when protective factors are not coded into the model. In this way, I am able to trace the political and ideological threads underpinning categorisation systems whose existing architecture is being used to build datafied pre-emptive prompts around life events while incorporating and retaining the power biases of those who designed them and mutually reinforcing gendered expectations. I unsettle the facticity of the risk categories that are generated by the child protection scoring model and am able to draw on findings about the criminalisation of care experienced girls which means that they are overrepresented in statutory databases. This would lead to data led approaches to pre-emptively identifying signatures of abuse and exploitation reifying conditions of their vulnerability, disenfranchisement and risk.

I am able to open up discussions about intersectional disadvantage compounded by datafication using substantive findings from the social policy literature that add a much needed depth to discussions around the datafied welfare state as articulated by Media and Communication Scholars (Niklas 2016, Eubanks 2018, Peña Gangadharan and Niklas 2019, Andrejevic et al. 2020, Kaun et al. 2023). There has been no research to date that employs this kind of deep interdisciplinary exploration of data driven systems even as an impressive body of work looking at ‘data justice’ and structural inequalities in automated welfare bureaucracies (Dencik et al. 2019, Dencik et al. 2022) has recently begun to accumulate.

This thesis has produced a detailed look at the advance of datafication in sensitive policy fields and it’s subsequent tensions by foregrounding the actual practices surrounding these data technologies.

Future research should be aimed at creating a longitudinal evidence base to understand if those who were targeted for early interventions through analytics and datafied modelling go on to do well in life. My research raised concerns about risk scoring through digital proxies as drawing on a deficit model and not being able to code in protective factors. Support systems that mitigate the effects of adverse childhood experiences should be coded into these data systems so that their outputs don't target those with who fit the ideal type of a vulnerable victim with no concomitant offer of support.

Future studies could explore how social workers and frontline staff navigate these systems by comparing a) child protection departments with in-house analytics teams that design these systems and have the ability to modify features and b) child protection departments using off the shelf platforms like OneView that is managed by external consultants. Frontline workers and their knowledge is what glues the system together. This could test the hypothesis of whether empowering the frontline social work staff member and creating organisational conditions where they don't feel stupid for not being statistical; and are able to speak up and ask for modifications from the data science team produces better outcomes. A systematic study of how these tools affect direct work with families and whether these are used blindly is much needed. During the course of this research, several councils abandoned algorithm based child protection - lessons learnt from these case studies and an investigation into whether these risk profiling models made it back in through the back door (within arrears and revenue collection sections of the department) are also vital to conduct.

Appendix 1

Topic guide

Background Information

1. Can you tell me a little bit about your role? How long have you been in it?

2. Can you tell me a bit more about the team that you lead? What is its scope? And what are the kinds of digital products that you work on?

3. Thinking back to the last few projects you worked on, what change or digital transformation challenge was being addressed?

4. Could you tell me about how you gathered the problem definition?

5. How did you interact with frontline workers/operational staff and convert their tasks into a digital process?

Uses of data analytics

6. Which team, project or good use cases come to mind where analytics or data is used particularly well to learn things about users?
    a. Do you use Google Analytics?

7. Could you take me through how you use data from the service to get granular, real time insights to understand patterns and risks? (from the Department Business Strategy)

8. One of the goals referred to in several of your department's policy documents was to 'develop analytical models and embed them into operational systems'. Could you explain how you achieve this?

Role of Data Scientists in the department

9. Could you take me through the role of data scientists in the department - how are they different to software engineers?

10. What are some of the challenges to implementing data driven solutions in an organisation that already has a certain way of doing things?

11. How do you interact with policy colleagues?

Appendix 2

Anonymised in depth Interviewee list

Case Study		**Interviewees**	**Date** **Contact or Follow up**	Transcript - Number of words
CASE STUDY 1 Council	6	Social Worker in children's social care	26-12-2020 Spoke several times	6880
Council Off the record	7	Senior Manager, Transformation	21-06-2021 Once	5491
Council	8	Manager, Early Intervention and Safer Communities	21-06-2021 Spoke four times	10468
Council	9	Analytics Hub Manager	2-07-2021 Once	12028
Council	10	Former Data Scientist for council	6-07-2021	10260
Department A	11	Data Scientist and manager	8-07-2021	10155
Department A	12	Digital performance analyst	9-07-2021	
Department A Wants to review Transcript	13	Head of Role for Data Science	9-07-2021 Spoke twice	9127
Council	14.	Business Intelligence Developer	18-8-2021 Spoke twice	11885
Council	15.	Education Inclusion Manager	23-08-2021 Once	9088
Council	16.	Education Inclusion Manager	23-08-2021 Once	6751
Council	17.	Safer Options Manager	24-08-2021 Once	6243
Department A	18.	Former assistant to IT professions. Now is an academic and consultant	1-9-2021 Spoke four times	11900 + 12495

		with the Council of Europe, working locally and internationally on a range of AI issues		
Department A	19.	Former Policy Expert – worked with analytics based service redesign consultancy	20-11-2021	7049
Department A	20.	Former Digital Government civil servant, Policy Expert attached to Number 10	26-11-2021	8227
Case Study 2 Department A	1	Senior Data Scientist	28-09-2020 Spoke twice	17081
Department A	2	Senior delivery manager	19-10-2020 Spoke three times and email contact	14238
Department A	3	A User Researcher	20-10-2020 Spoke once and detailed emails reponding to further questions	3901
Department A Off the record	4	Product Manager	30-10-2020 Spoke twice	5677
Department A Wants to review Transcript	5	Lead Software Engineer	10-11-2020 Spoke once	6058

References

Adams, V. (2016). Metrics: What Counts in Global Health, Duke University Press.

Alkhatib, A. and M. Bernstein (2019). Street-Level Algorithms: A Theory at the Gaps between Policy and Decisions. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, ACM.

Alston, P. (2019). Report of the Special Rapporteur on Extreme Poverty and Human Rights (Un Doc. A/74/48037), United Nations.

Amoore, L. (2013). The Politics of Possibility: Risk and Security Beyond Probability, Duke University Press.

Amoore, L. (2020). Cloud Ethics: Algorithms and the Attributes of Ourselves and Others, Duke University Press.

Amoore, L. (2020). Why 'Ditch the Algorithm'is the Future of Political Protest. The Guardian **19**

Ananny, M. (2016). Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness. Science, Technology, & Human Values **41**(1): 93-117

Anderson, R., I. Brown, R. Clayton, T. Dowty, D. Korff and E. Munro (2006). Children's Databases – Safety and Privacy. A Report for the Information Commissioner. Information Commissioner's Office: Wilmslow, Foundation for Information Policy Research

Andrejevic, M., L. Dencik and E. Treré (2020). From Pre-Emption to Slowness: Assessing the Contrasting Temporalities of Data-Driven Predictive Policing. New Media & Society **22**(9): 1528-1544. 10.1177/1461444820913565

Attride-Stirling, J. (2001). Thematic Networks: An Analytic Tool for Qualitative Research. Qualitative research **1**(3): 385-405

Authority, F. C. (2020). The Financial Lives of Consumers across the Uk. Key Findings from the Fca's Financial Lives Survey 2017. F. C. A. Strategy and Competition Division.

Baack, S. (2015). Datafication and Empowerment: How the Open Data Movement Re-Articulates Notions of Democracy, Participation, and Journalism. Big Data & Society **2**(2): 2053951715594634

Banyard, V., S. Hamby and J. Grych (2017). Health Effects of Adverse Childhood Events: Identifying Promising Protective Factors at the Intersection of Mental and Physical Well-Being. Child Abuse & Neglect **65**: 88-98. 10.1016/j.chiabu.2017.01.011

Barnes, A. and J. Parkhurst (2014). Can Global Health Policy Be Depoliticized? A Critique of Global Calls for Evidence-Based Policy. Handbook of Global Health Policy. UK: John Wiley & Sons, Ltd

Barocas, S., K. Crawford, A. Shapiro and H. Wallach (2017). The Problem with Bias: From Allocative to Representational Harms in Machine Learning. SIGCIS conference paper.

Barocas, S. and A. D. Selbst (2016). Big Data's Disparate Impact. Cal. L. Rev. **104**: 671

Bauer, M. W. and G. Gaskell (2000). Qualitative Researching with Text, Image and Sound: A Practical Handbook for Social Research. London, Sage.

Beaussier, A.-L., D. Demeritt, A. Griffiths and H. Rothstein (2016). Accounting for Failure: Risk-Based Regulation and the Problems of Ensuring Healthcare Quality in the Nhs. Health, risk and society **18**(3-4): 205-224

Bechmann, A. and G. Bowker (2019). Unsupervised by Any Other Name: Hidden Layers of Knowledge Production in Artificial Intelligence on Social Media. Big Data and Society **6**(1)

Becker, H. and B. Geer (1957). Participant Observation and Interviewing: A Comparison. Human organization **16**(3): 28-32

Beer, D. (2016). How Should We Do the History of Big Data? Big Data & Society **3**(1): 2053951716646135. 10.1177/2053951716646135

Bennett, A. and A. George (2005). Case Studies and Theory Development in the Social Sciences. Cambridge, Massachussetts, MIT Press.

Bennett Moses, L. and L. d. Koker (2017). Open Secrets: Balancing Operational Secrecy and Transparency in the Collection and Use of Data by National Security and Law Enforcement Agencies. Melb. UL Rev. **41**: 530

Beresford, P. (2001). Evidence-Based Care: Developing the Discussion. Journal of Integrated Care

Berger, A. A. (1998). Media Research Techniques, Sage Publications.

Binns, R. (2019). Human Judgement in Algorithmic Loops; Individual Justice and Automated Decision-Making In. Available at SSRN: https://ssrn.com/abstract=3452030 or http://dx.doi.org/10.2139/ssrn.3452030.

Bird, S., S. Barocas, K. Crawford, F. Diaz and H. Wallach (2016). Exploring or Exploiting? Social and Ethical Implications of Autonomous Experimentation in Ai. Workshop on Fairness, Accountability, and Transparency in Machine Learning.

Blank, G. (2020). Who Uses E-Government in Britain? A Report to the Department for Digital, Culture, Media and Sport.

Boaz, A., L. Grayson, R. Levitt and W. Solesbury (2008). Does Evidence-Based Policy Work? Learning from the Uk Experience. Evidence & Policy: A Journal of Research, Debate and Practice **4**(2): 233-253

Bolukbasi, T., K.-W. Chang, J. Y. Zou, V. Saligrama and A. T. Kalai (2016). Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings. Advances in neural information processing systems.

Bouk, D. (2015). How Our Days Became Numbered. How Our Days Became Numbered In, University of Chicago Press.

Bowker, G. C. and S. L. Star (2000). Sorting Things Out: Classification and Its Consequences, MIT press.

Boyd, D. and K. Crawford (2012). Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. Information, communication & society **15**(5): 662-679

Braun, V. and V. Clarke (2006). Using Thematic Analysis in Psychology. Qualitative research in psychology **3**(2): 77-101

Broussard, M. (2023). More Than a Glitch: Confronting Race, Gender, and Ability Bias in Tech, MIT Press.

Bryman, A. (2013). Introduction: 'Inside' Accounts and Social Research in Organizations. Doing Research in Organizations (Rle: Organizations) In, Routledge**:** 11-30.

Burman, M. and L. Gelsthorpe (2017). Feminist Criminology: Inequalities, Powerlessness, and Justice.

Burrell, J. (2016). How the Machine ‘Thinks’: Understanding Opacity in Machine Learning Algorithms. Big Data and Society **3**(1): 2053951715622512

Cabinet Office (1999). Modernising Government : Action Plan, London: Cabinet Office.

Cairney, P. (2019). The Uk Government’s Imaginative Use of Evidence to Make Policy. British Politics **14**(1): 1-22

Cameron, D. (2011). Pm’s Speech on the Fightback after the Riots, Https://Www.Gov.Uk/Government/Speeches/Pms-Speech-on-the-Fightback-after-the-Riots Accessed 3rd March 2023.

Canguilhem, G. (1989). The Normal and the Pathological. New York: Zone Book.

Carney, T. (2018). The New Digital Future for Welfare: Debts without Legal Proofs or Moral Authority? UNSW Law Journal Forum

Carter, L. (2021). Prescripted Living: Gender Stereotypes and Data-Based Surveillance in the Uk Welfare State. Internet Policy Review **10**(4): 1-23

Cawson, P., C. Wattam, S. Brooker and G. Kelly (2000). Child Maltreatment in the United Kingdom: A Study of the Prevalence of Child Abuse and Neglect. London, NSPCC.

Cheney-Lippold, J. (2011). A New Algorithmic Identity: Soft Biopolitics and the Modulation of Control. Theory, Culture & Society **28**(6): 164-181

Cheney-Lippold, J. (2017). We Are Data: Algorithms and the Making of Our Digital Selves, NYU Press.

Christin, A. (2017). Algorithms in Practice: Comparing Web Journalism and Criminal Justice. Big Data & Society **4**(2): 2053951717718855

Christin, A. (2020). Metrics at Work: Journalism and the Contested Meaning of Algorithms, Princeton University Press.

Citron, D. K. and F. Pasquale (2014). The Scored Society: Due Process for Automated Predictions. Wash. UL Rev. **89**: 1

Cohen, J. E. (2015). The Biopolitical Public Domain.

Community, T. T. W. (2022). The Turing Way: A Handbook for Reproducible, Ethical and Collaborative Research. Zenodo. doi: 10.5281/zenodo.3233853

Couldry, N. (2013). A Necessary Disenchantment: Myth, Agency and Injustice in the Digital Age. Sociological Review

Couldry, N. and A. Hepp (2016). The Mediated Construction of Reality, John Wiley & Sons.

Couldry, N. and U. Mejias (2019). The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism. Stanford, California: Stanford University Press.

Couldry, N. and U. A. Mejias (2019). Data Colonialism: Rethinking Big Data’s Relation to the Contemporary Subject. Television & New Media **20**(4): 336-349

Crawford, K. (2013). The Hidden Biases in Big Data. Harvard Business Review **1**(4)

Crawford, K. (2014). The Test We Can—and Should—Run on Facebook. The Atlantic **2**

Crawford, K. (2016). Can an Algorithm Be Agonistic? Ten Scenes from Life in Calculated Publics. Science, Technology, & Human Values **41**(1): 77-92

Crooks, R. (2019). Cat-and-Mouse Games: Dataveillance and Performativity in Urban Schools. Surveillance & Society **17**(3/4): 484-498

Cuccaro-Alamin, S., R. Foust, R. Vaithianathan and E. Putnam-Hornstein (2017). Risk Assessment and Decision Making in Child Protective Services: Predictive Risk Modeling in Context. Children and Youth Services Review **79**: 291-298

Cummings, D. (2020). 'Two Hands Are a Lot' — We're Hiring Data Scientists, Project Managers, Policy Experts, Assorted Weirdos….Dominic Cummings's Blog Retrieved from. https://dominiccummings.com/2020/01/02/two-hands-are-a-lot-were-hiring-data-scientists-project-managers-policy-experts-assorted-weirdos/.

Currie, M. and W. U. Hsu (2019). Performative Data: Cultures of Government Data Practice. Journal of Cultural Analytics **4**(1)

Currie, M., J. Knox and C. Mcgregor (2023). Data Justice and the Right to the City, Edinburgh University Press.

D'ignazio, C. and L. F. Klein (2020). Data Feminism, MIT press.

DCMS (2020). National Data Strategy. Department for Digital, Culture, Media and Sport

Dean, J. P. and W. F. Whyte (1958). How Do You Know If the Informant Is Telling the Truth? Human organization **17**(2): 34

defenddigitalme (2020). The State of Data 2020: Mapping a Child's Digital Footprint in the State Education Landscape in England.

Dencik, L., A. Hintz, J. Redden and E. Treré (2019). Exploring Data Justice: Conceptions, Applications and Directions, Taylor & Francis. **22:** 873-881.

Dencik, L., A. Hintz, J. Redden and E. Treré (2019). Exploring Data Justice: Conceptions, Applications and Directions. Information, communication & society **22**(7): 873-881. 10.1080/1369118X.2019.1606268

Dencik, L., A. Hintz, J. Redden and E. Treré (2022). Data Justice, SAGE.

Dencik, L., A. Hintz, J. Redden and H. Warne (2018). Data Scores as Governance: Investigating Uses of Citizen Scoring in Public Services Project Report.

Desrosières, A. (1998). The Politics of Large Numbers: A History of Statistical Reasoning, Harvard University Press.

Deutscher, I. (1973). What We Say/What We Do: Sentiments & Acts. Glenview, Illinois, Scott, Foresman.

Diakopoulos, N. (2014). Algorithmic Accountability Reporting: On the Investigation of Black Boxes.

Diakopoulos, N. and M. J. D. J. Koliska (2017). Algorithmic Transparency in the News Media. **5**(7): 809-828

Dickinson, H. and J. Glasby (2010). The Personalisation Agenda: Implications for the Third Sector.

DiMaggio, P. (1997). Culture and Cognition. Annual review of sociology **23**(1): 263-287

Dodge, M. and R. Kitchin (2005). Code and the Transduction of Space. Annals of the Association of American geographers **95**(1): 162-180

Douglas, M. (1986). How Institutions Think. Syracuse, New York, Syracuse University Press.

Dourish, P. (2016). Algorithms and Their Others: Algorithmic Culture in Context. Big Data & Society **3**(2)

Dunleavy, P. and H. Margetts (2015). Design Principles for Essentially Digital Governance.

DWP (2017). Improving Lives: Helping Workless Families, Https://Www.Gov.Uk/Government/Uploads/System/Uploads/Attachment_Data/File/621364/Improving-Lives-Helping-Workless-Families-Web-Version.Pdf. London, Department for Work and Pensions.

Edwards, R., V. Gillies and S. Gorin (2022). Problem-Solving for Problem-Solving: Data Analytics to Identify Families for Service Intervention. Critical Social Policy **42**(2): 265-284

Engelhardt, T. (2013). The Us Has Bombed at Least Eight Wedding Parties since 2001 Why Are So Many of Our “Surgical” Strikes Going So Terribly Wrong? The Nation.

Ernst & Young (2019). How Can Data Tell a Story That Keeps a Vulnerable Person Safe? Safeguarding Vulnerable People to Improve Wellbeing in a Sustainable Way.

Eubanks, V. (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. . New York, St. Martin’s Press.

Ezzy, D. (2013). Qualitative Analysis: Practice and Innovation. London and New York, Routledge.

Fabijan, A., P. Dmitriev, H. H. Olsson and J. Bosch (2018). Online Controlled Experimentation at Scale: An Empirical Survey on the Current State of a/B Testing. 2018 44th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), IEEE.

Ferguson, I. (2007). Increasing User Choice or Privatizing Risk? The Antinomies of Personalization. British journal of social work **37**(3): 387-403

Fiore-Silfvast, B. and G. Neff (2013). What We Talk About When We Talk Data: Valences and the Social Performance of Multiple Metrics in Digital Health. Ethnographic Praxis in Industry Conference Proceedings, Wiley Online Library.

Fischer, F. (2003). Reframing Public Policy: Discursive Politics and Deliberative Practices. Oxford, Oxford University Press.

Fischer, F. (2015). In Pursuit of Usable Knowledge: Critical Policy Analysis and the Argumentative Turn. Handbook of Critical Policy Studies In D. T. Frank Fischer and M. O. Anna Durnová (Eds.). Cheltenham, UK; Northampton, MA, USA Edward Elgar Publishing.

Fischer, F., D. Torgerson, A. Durnová and M. Orsini, Eds. (2015). Handbook of Critical Policy Studies. Cheltenham, UK; Northampton, MA, USA Edward Elgar Publishing.

Fitzpatrick, C. (2020). Reconsidering the Care-Crime Connection in a Climate of Crisis. Child and Family Law Quarterly **32**(2): 103-118

Fitzpatrick, C. (2022). Challenging Perceptions of Care-Experienced Girls and Women. Prison Service Journal(258): 19-24

Fitzpatrick, C., K. Hunter, J. Shaw and J. Staines (2022). Disrupting the Routes between Care and Custody for Girls and Women. Lancaster University & The Nuffield Foundation.

Florescu, D., M. Karlberg, F. Reis, P. R. Del Castillo, M. Skaliotis and A. Wirthmann (2014). Will 'Big Data'transform Official Statistics. European Conference on the QualityofOfficial Statistics. Vienna, Austria.

Floridi, L. (2011). The Informational Nature of Personal Identity. Minds and Machines **21**(4): 549

Floridi, L. (2012). Big Data and Their Epistemological Challenge. Philosophy and Technology **25**(4): 435-437

Floridi, L. (2014). The Fourth Revolution: How the Infosphere Is Reshaping Human Reality, OUP Oxford.

Flyvberg, B. (2011). Case Study. The Sage Handbook of Qualitative Research, 4th Edition In Norman K. Denzin and Y. S. Lincoln (Eds.). Thousand Oaks, CA, Sage**:** 301-316.

Force;, S. E. T. (2007). Families at Risk. London, Cabinet Office.

Fourcade, M. and K. Healy (2013). Classification Situations: Life-Chances in the Neoliberal Era. Accounting, Organizations and Society **38**(8): 559-572

Fredman, S. (1998). Women and the Law, Clarendon Press ; Oxford University Press.

Freud, D. (2021). Clashing Agendas: Inside the Welfare Trap, Nine Elms Books.

Fussey, P. and A. Sandhu (2022). Surveillance Arbitration in the Era of Digital Policing. Theoretical criminology **26**(1): 3-22

Garvin, T. and J. Eyles (2001). Public Health Responses for Skin Cancer Prevention: The Policy Framing of Sun Safety in Australia, Canada and England. Social Science Medicine **53**(9): 1175-1189

Gillespie, T. (2014). The Relevance of Algorithms. Media technologies: Essays on communication, materiality, and society **167**(2014): 167

Gillingham, P. (2015). Predictive Risk Modelling to Prevent Child Maltreatment and Other Adverse Outcomes for Service Users: Inside the 'Black Box'of Machine Learning. The British Journal of Social Work **46**(4): 1044-1058

Gillingham, P. (2016). Predictive Risk Modelling to Prevent Child Maltreatment and Other Adverse Outcomes for Service Users: Inside the 'Black Box'of Machine Learning. The British Journal of Social Work **46**(4): 1044-1058

Gillingham, P. (2019). The Development of Algorithmically Based Decision-Making Systems in Children's Protective Services: Is Administrative Data Good Enough? The British Journal of Social Work **50**(2): 565-580. 10.1093/bjsw/bcz157

Gitelman, L. (2013). Raw Data Is an Oxymoron, MIT press.

Glasby, J. and R. Littlechild (2009). Direct Payments and Personal Budgets: Putting Personalisation into Practice, Policy Press.

Grabham, E. and J. Smith (2010). From Social Security to Individual Responsibility (Part Two): Writing Off Poor Women's Work in the Welfare Reform Act 2009. Journal of Social Welfare & Family Law **32**(1): 81-93

Gray, M., D. Plath and S. Webb (2009). Evidence-Based Social Work: A Critical Stance, Routledge.

Green, B. and Y. Chen (2019). Disparate Interactions: An Algorithm-in-the-Loop Analysis of Fairness in Risk Assessments. ACM Conference on Fairness, Accountability, and Transparency.

Greenland, C. (1987). Preventing Can Deaths: An International Study of Deaths Due to Child Abuse and Neglect, Tavistock/Routledge.

Griffiths, H. (2000). Reflections on Risk in Child Protection: Is It Time for a Shift in Paradigms? Child & Family Social Work **5**(1): 1-10. https://doi.org/10.1046/j.1365-2206.2000.00145.x

Grill, G. (2022). Constructing Certainty in Machine Learning: On the Performativity of Testing and Its Hold on the Future.

Grimmelmann, J. (2015). The Law and Ethics of Experiments on Social Media Users. Colo. Tech. LJ **13**: 219

Guest, G., K. M. MacQueen and E. E. Namey (2011). Applied Thematic Analysis. Thousand Oaks, California, Sage Publications.

Hacking, I. (1982). Biopower and the Avalanche of Printed Numbers. Humanities in society **5**(3-4): 279-295

Hacking, I. (1986). Making up People.

Hall, C., S. Slembrouck and S. Sarangi (2006). Language Practices in Social Work: Categorisation and Accountability in Child Welfare, Psychology Press.

Hammersley, M. (2012). What Is Qualitative Research? London and New York, A&C Black.

Hayden, C. and C. Jenkins (2014). 'Troubled Families' Programme in England:'Wicked Problems' and Policy-Based Evidence. Policy Studies **35**(6): 631-649

Head, B. W. (2008). Three Lenses of Evidence-Based Policy. Australian Journal of Public Administration **67**(1): 1-11

Heeks, R. (2005). Implementing and Managing Egovernment: An International Text, Sage.

Heemsbergen, L., E. Treré and G. Pereira (2022). Introduction to Algorithmic Antagonisms: Resistance, Reconfiguration, and Renaissance for Computational Life. Media International Australia **183**(1): 3-15

Henman, P. and M. Adler (2003). Information Technology and the Governance of Social Security. Critical Social Policy **23**(2): 139-164

Hildebrandt, M. (2011). Who Needs Stories If You Can Get the Data? Isps in the Era of Big Number Crunching. Philosophy & Technology **24**(4): 371-390

Hildebrandt, M. and B. J. Koops (2010). The Challenges of Ambient Law and Legal Protection in the Profiling Era. The Modern Law Review **73**(3): 428-460

Hjörne, E. and A.-C. Evaldsson (2016). Disability Identities and Category Work in Institutional Practices. The Routledge Handbook of Language and Identity: 396

Hjörne, E. and R. Säljö (2004). The Pupil Welfare Team as a Discourse Community: Accounting for School Problems. Linguistics and education **15**(4): 321-338

HOC (2019). Consumers' Access to Financial Services. Increasing Financial Inclusion. . H. o. C. T. S. Committee.

Hoffmann, A. L. (2021). Terms of Inclusion: Data, Discourse, Violence. New Media & Society **23**(12): 3539-3556

Hogan, B. (2021). Networks Are a Lens for Power: A Commentary on the Recent Advances in the Ethics of Social Networks Special Issue. Social Networks **67**: 9-12

Hunter, S. (2016). Feminist Perspectives. The Student's Companion to Social Policy (Fifth Edition) In T. H. P. Alcock, M. May, & S. Wright (Ed.). Hoboken, NJ, John Wiley & Sons.

Institute, A. L. (2020). Transparency Mechanisms for Uk Public-Sector Algorithmic Decision-Making Systems. Explainer for Government, Local Government, Policymakers and Researchers.

Institute, A. L., A. N. Institute and O. G. Partnership (2021). Algorithmic Accountability for the Public Sector.

Javid, S. (2021). Sajid Javid's Speech to Conservative Party Conference in Full. https://www.conservatives.com/news/2021/sajid-javid-s-speech-to-conservative-party-conference.

John, P. (2018). How Far to Nudge?: Assessing Behavioural Public Policy, Edward Elgar Publishing.

Jones, H. (2013). Negotiating Cohesion, Inequality and Change: Uncomfortable Positions in Local Government. Bristol, The Policy Press.

Justice, C. f. S. (2007). Breakthrough Britain. London, Centre for Social Justice.

Justice, C. f. S. (2013). Fractured Families. London, Centre for Social Justice.

Justice, C. f. W. s. (2020). Super-Complaint by Centre for Women’s Justice: Failure to Address Police Perpetrated Domestic Abuse. London, Centre for Women’s Justice.

Kaun, A., S. Lomborg, C. Pentzold, D. Allhutter and K. Sztandar-Sztanderska (2023). Crosscurrents: Welfare. Media, Culture & Society: 01634437231154777

Kay, A. (2011). Evidence-Based Policy-Making: The Elusive Search for Rational Public Administration. Australian Journal of Public Administration **70**(3): 236-245

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

Keddell, E. (2019). Algorithmic Justice in Child Protection: Statistical Fairness, Social Justice and the Implications for Practice. Social Sciences **8**(10): 281

Kendall, S., J. Rodger and H. Palmer (2010). Redesigning Provision for Families with Multiple Problems: An Assessment of the Early Impact of Different Local Approaches, Department for Education.

Kennedy, H., T. Poell and J. van Dijck (2015). Data and Agency. Big Data & Society **2**(2): 1-7

Kennedy, H., M. Taylor, S. Oman, J. Bates, I. Medina-Perea, H. a. Ditchfield and L. Pinney (2021). Context and Prior Awareness Matter in Public Concerns About Data Uses: Findings from the 2020 Living with Data Survey, Https://Livingwithdata.Org/Project/Wp-Content/Uploads/2021/07/Living-with-Data-2020-Survey-Summary-Report.Pdf. (accessed March 2023)

Kitchin, R. (2014). Conceptualising Data.

Kitchin, R. (2017). Thinking Critically About and Researching Algorithms. Information, Communication and Society **20**(1): 14-29

Kitchin, R. and T. Lauriault (2014). Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work. The Programmable City Working Paper 2.

Kitchin, R. and G. McArdle (2016). What Makes Big Data, Big Data? Exploring the Ontological Characteristics of 26 Datasets. Big Data & Society **3**(1): 2053951716631130

Kohavi, R., A. Deng, B. Frasca, T. Walker, Y. Xu and N. Pohlmann (2013). Online Controlled Experiments at Large Scale. Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining.

Kolkman, D. (2022). The (in)Credibility of Algorithmic Models to Non-Experts. Information, communication & society **25**(1): 93-109. 10.1080/1369118X.2020.1761860

Kolkman, D. A., P. Campo, T. Balke-Visser and N. Gilbert (2016). How to Build Models for Government: Criteria Driving Model Acceptance in Policymaking. Policy Sciences **49**: 489-504

König, P. D. (2017). The Place of Conditionality and Individual Responsibility in a "Data-Driven Economy". Big Data & Society **4**(2): 2053951717742419

Kraemer, K. L. and J. L. King (1986). Computer-Based Systems for Cooperative Work and Group Decisionmaking: Status of Use and Problems in Development. Proceedings of the 1986 ACM conference on Computer-supported cooperative work, ACM.

Kramer, A. D., J. E. Guillory and J. T. Hancock (2014). Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks. Proceedings of the National Academy of Sciences **111**(24): 8788-8790

Kvale, S. (1996). Interviews: An Introduction to Qualitative Research Interviewing. Thousand Oaks, Sage.

Lambert, M. (2019). Between "Families in Trouble" and "Children at Risk": Historicising "Troubled Family" Policy in England since 1945. Children & Society **33**(1): 82-91

Leadbeater, C. (2004). Personalisation through Participation: A New Script for Public Services, Demos.

Leese, M. (2014). The New Profiling: Algorithms, Black Boxes, and the Failure of Anti-Discriminatory Safeguards in the European Union. Security Dialogue **45**(5): 494-511

Lewis, J., Ed. (2000). Family Policy in the Post-War Period. Cross-Currents: Family Law and Policy in the United States and England. Oxford, Oxford University Press.

Lewis, J. (2001). What Works in Community Care? Journal of Integrated Care **9**(1)

Lewis, J. and H. Glennerster (1996). Implementing the New Community Care, Open University Press.

Lilleker, D. (2003). Interviewing the Political Elite: Navigating a Potential Minefield. Politics **23**(3): 207-214

Lipsky, M. (1980). Street-Level Bureaucracy: Dilemmas of the Individual in Public Service, Russell Sage Foundation.

Lupton, D. (2015). The Thirteen Ps of Big Data. This Sociological Life

Lyle, C. G. and E. Graham (2000). Looks Can Be Deceiving: Using a Risk Assessment Instrument to Evaluate the out-Comes of Child Protection Services. Children Youth Services Review **22**(11-12): 935-949

Majone, G. (1989). Evidence, Argument, and Persuasion in the Policy Process. New Haven Yale University Press.

Malomo, F. and V. Sena (2017). Data Intelligence for Local Government? Assessing the Benefits and Barriers to Use of Big Data in the Public Sector. Policy & Internet **9**(1): 7-27

Malone, T. W., J. Yates and R. I. Benjamin (1994). Electronic Markets and Electronic Hierarchies, New York: Oxford University Press.

Margetts, H. and P. Dunleavy (2013). The Second Wave of Digital-Era Governance: A Quasi-Paradigm for Government on the Web. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **371**(1987): 20120382

Markson, T. (2022). Gds Boss Puts 'Mobile-First and Hyperpersonalisation' at Heart of Future Service Transformation. Public Technology. https://www.publictechnology.net/articles/news/gds-boss-puts-%E2%80%98mobile-first-and-hyperpersonalisation%E2%80%99-heart-future-service.

Marr, B. (2014). Big Data: The 5 Vs Everyone Must Know. March 6.

Martin W Bauer and George Gaskell (2000). Individual and Group Interviewing: Qualitative Researching with Text, Image and Sound. London SAGE Publications Ltd.

Marz, N. and J. Warren (2015). Big Data: Principles and Best Practices of Scalable Realtime Data Systems, Manning Publications Co.

Matias, J. N. (2017). Governing Human and Machine Behavior in an Experimenting Society, Massachusetts Institute of Technology.

Mayer-Schönberger, V. and K. Cukier (2013). Big Data: A Revolution That Will Transform How We Live, Work, and Think. Houghton Mifflin Harcourt

McGovern, P. (2012). Hrm, Technical Workers and the Multinational Corporation. Hoboken Taylor and Francis.

McGovern, P. and D. Alburez-Gutierrez (2017). Who Takes Workplace Case-Study Methods Seriously? The Influence of Gender, Academic Rank and Phd Training. Industrial Relations Journal **48**(2): 98-114

McNeill, G., J. Bright and S. A. Hale (2017). Estimating Local Commuting Patterns from Geolocated Twitter Data. EPJ Data Science **6**(1): 24. 10.1140/epjds/s13688-017-0120-x

McNulty, E. (2014). Understanding Big Data: The Seven V's. Dataconomy, May **22**

McQuillan, D. (2018). Data Science as Machinic Neoplatonism. Philosophy & Technology **31**(2): 253-272

Metcalf, J. and K. Crawford (2016). Where Are Human Subjects in Big Data Research? The Emerging Ethics Divide. Big Data & Society **3**(1): 2053951716650211

Meyer, M. N. (2015). Two Cheers for Corporate Experimentation: The a/B Illusion and the Virtues of Data-Driven Innovation. Colo. Tech. LJ **13**: 273

Meyer, M. N. and C. F. Chabris (2015). Please, Corporations, Experiment on Us. The New York Times **19**

Mittelstadt, B. D., P. Allo, M. Taddeo, S. Wachter and L. Floridi (2016). The Ethics of Algorithms: Mapping the Debate. Big Data and Society **3**(2): 2053951716679679

Mohr, J. W. and V. Duquenne (1997). The Duality of Culture and Practice: Poverty Relief in New York City, 1888--1917. Theory and society **26**(2-3): 305-356

Morland, P. (2022). Should We Tax the Childless? The Times. London, https://www.thetimes.co.uk/article/should-we-tax-the-childless-j7h9c297r.

Munro, E. (2019). Predictive Analytics in Child Protection. CHESS Working Paper No. 2019-03 Produced as part of the Knowledge for Use (K4U) Research Project], Centre for Humanities Engaging Science and Society (CHESS), Durham University.

Munro, E., N. Cartwright, J. Hardie and E. Montuschi (2017). Improving Child Safety: Deliberation, Judgement and Empirical Research, Centre for Humanities Engaging Science and Society (CHESS).

Murray, D., P. Fussey, L. McGregor and M. Sunkin (2020). Effective Oversight of Large-Scale Surveillance Activities: A Human Rights Perspective. Journal of National Security Law and Policy **11**: 743

Needham, C. (2011). Personalising Public Services: Understanding the Personalisation Narrative, Policy Press.

Neff, G., A. Tanweer, B. Fiore-Gartland and L. Osburn (2017). Critique and Contribute: A Practice-Based Framework for Improving Critical Data Studies and Data Science. Big data **5**(2): 85-97

Newell, S. and M. Marabelli (2015). Strategic Opportunities (and Challenges) of Algorithmic Decision-Making: A Call for Action on the Long-Term Societal Effects of ‘Datification’. The Journal of Strategic Information Systems **24**(1): 3-14

Niklas, J. (2016). E-Government in the Welfare State–Human Rights Implications of Digitization of Social Policy in Poland. Global Information Society Watch: 182-188

Nissenbaum, H. (2009). Privacy in Context: Technology, Policy, and the Integrity of Social Life. Stanford, California, Stanford University Press.

Noble, S. U. (2018). Algorithms of Oppression. Algorithms of Oppression In, New York University Press.

Noble, S. U. (2018). Algorithms of Oppression: How Search Engines Reinforce Racism, NYU Press.

Noble, V. (2008). Inside the Welfare State: Foundations of Policy and Practice in Post-War Britain, Routledge.

O'Neil, C. (2017). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy, Broadway Books.

Oak, E. (2015). A Minority Report for Social Work? The Predictive Risk Model (Prm) and the Tuituia Assessment Framework in Addressing the Needs of New Zealand's Vulnerable Children. The British Journal of Social Work **46**(5): 1208-1223

Oak, E. (2016). A Minority Report for Social Work? The Predictive Risk Model (Prm) and the Tuituia Assessment Framework in Addressing the Needs of New Zealand's Vulnerable Children. The British Journal of Social Work **46**(5): 1208-1223

OFCOM (2022). Adults’ Media Use and Attitudes Report 2022.

OFCOM (2022). The Communications Market 2022: Online Nation.

Orlikowski, W. J. (1989). Division among the Ranks: The Social Implications of Case Tools for System Developers.

Orlikowski, W. J. and J. Baroudi (1991). Studying Information Technology in Organizations: Research Approaches and Assumptions. Information Systems Research **2**(1): 1-28

Østergaard Møller, M. and G. Sommer Harrits (2013). Constructing at-Risk Target Groups. Critical Policy Studies **7**(2): 155-176

Oswald, M. (2022). A Three-Pillar Approach to Achieving Trustworthy and Accountable Use of Ai and Emerging Technology in Policing in England and Wales: Lessons from the West Midlands Data Ethics Model. European Journal of Law and Technology **13**(1)

Panel, R. C. a. V. (2012). After the Riots: Final Report of the Riots Communities and Victims Panel.

Parkhurst, J. (2017). The Politics of Evidence: From Evidence-Based Policy to the Good Governance of Evidence. Abingdon, Oxon, UK, Routledge.

Parsons, W. (2017). Policy Analysis in Britain. Handbook of Public Policy Analysis: Theory, Politics, and Methods In F. Fischer, G. J. Miller and M. S. Sidney (Eds.). New York London, Routledge**:** 563-578.

Pasquale, F. (2015). The Black Box Society: The Secret Algorithms That Control Money and Information, Harvard University Press.

Peña Gangadharan, S. and J. Niklas (2019). Decentering Technology in Discourse on Discrimination. Information, communication & society **22**(7): 882-899

Petty, T., M. Saba, T. Lewis, S. P. Gangadharan and V. Eubanks (2018). Reclaiming Our Data: Interim Report. Detroit.

Pleace, N. (2007). Workless People and Surveillant Mashups: Social Policy and Data Sharing in the Uk. Information, communication & society **10**(6): 943-960. 10.1080/13691180701751171

Policing, C. o. (2022). Police Perpetrated Domestic Abuse: Report on the Centre for Women's Justice Super Complaint. College of Policing.

Pollitt, C. (2010). Technological Change: A Central yet Neglected Feature of Public Administration. NISPAcee Journal of Public Administration and Policy **3**(2): 31-53

Powell, M. (1999). New Labour, New Welfare State?: The'third Way'in British Social Policy, Policy Press.

Prior, L., D. Hughes and S. Peckham (2012). The Discursive Turn in Policy Analysis and the Validation of Policy Stories. Journal of Social Policy **41**(2): 271-289

Pybus, J., M. Coté and T. Blanke (2015). Hacking the Social Life of Big Data. Big Data & Society **2**(2): 2053951715616649

Redden, J. (2015). Big Data as System of Knowledge: Investigating Canadian Governance. Compromised Data: From Social Media to Big Data In Greg Elmer, Joanna Redden and G. Langlois (Eds.). London, Bloomsbury**:** 17.

Redden, J., J. Brand, I. Sander and H. Warne (2022). Datafied Child Welfare Services as Sites of Struggle. Data Justice and the Right to the City In M. Currie, Jeremy Knox and C. Mcgregor (Eds.). Edinburgh, Edinburgh University Press**:** 69.

Review;, H. G. P. (2007). Building on Progress: Public Services. P. M. s. S. Unit. London.

Rhodes, R. (2005). Everyday Life in a Ministry: Public Administration as Anthropology. The American Review of Public Administration **35**(1): 3-25

Rhodes, R. (2011). Everyday Life in British Government, Oxford University Press.

Richards, D. (1996). Elite Interviewing: Approaches and Pitfalls. Politics **16**(3): 199-204

Robertson, H. and J. Travaglia (2015). Big Data Problems We Face Today Can Be Traced to the Social Ordering Practices of the 19th Century. Impact of Social Sciences Blog

Rosenblat, A. and L. Stark (2016). Algorithmic Labor and Information Asymmetries: A Case Study of Uber's Drivers. International Journal of Communication **10**: 27

Rubin, H. J. and I. S. Rubin (2011). Qualitative Interviewing: The Art of Hearing Data. Thousand Oaks, California, Sage.

Ruppert, E., P. Harvey, C. Lury, A. Mackenzie, R. McNally, S. A. Baker, Y. Kallianos and C. Lewis (2015). Socialising Big Data: From Concept to Practice. CRESC Working Paper Series(138)

Sanderson, I. (2002). Making Sense of 'What Works': Evidence Based Policy Making as Instrumental Rationality? Public Policy and Administration **17**(3): 61-75

Sanderson, I. (2006). Complexity,'Practical Rationality'and Evidence-Based Policy Making. Policy & politics **34**(1): 115-132

Sandvig, C. (2018). You Are a Political Junkie and Felon Who Loves the Sound of Blenders: Machine Learning Taste Publics. International Communication Association 65th Annual Conference Caribe Hilton, San Juan, Puerto Rico

Sandvig;, C., K. Hamilton;, K. Karahalios; and C. Langbort; (2014). Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms. Data discrimination: converting critical concerns into productive inquiry **22**

Scourfield, P. (2005). Implementing the Community Care (Direct Payments) Act: Will the Supply of Personal Assistants Meet the Demand and at What Price? Journal of Social Policy **34**(3): 469-488

Seaver, N. (2017). Algorithms as Culture: Some Tactics for the Ethnography of Algorithmic Systems. Big Data & Society **4**(2): 2053951717738104

Shoshana, Z. (1988). In the Age of the Smart Machine: The Future of Work and Power. New York: Basic.

Simons, H. (2009). Case Study Research in Practice. London, SAGE publications.

Smith, K. (2006). Problematising Power Relations in 'Elite'interviews. Geoforum **37**(4): 643-653

Solesbury, W. (2002). The Ascendancy of Evidence. Planning Theory & Practice **3**(1): 90-96. 10.1080/14649350220117834

Stake, R. E. (2006). Multiple Case Study Analysis. New York, Guilford.

Stone, D. (2005). Foreword. Deserving and Entitled: Social Constructions and Public Policy In A. L. Schneider and H. M. Ingram (Eds.). Ithaca, SUNY Press.

Suresh, H. and J. V. Guttag (2019). A Framework for Understanding Unintended Consequences of Machine Learning. https://arxiv.org/abs/1901.10002

Sweeney, L. (2013). Discrimination in Online Ad Delivery. arXiv preprint arXiv:1301.6822

Tanweer, A. (2018). Data Science of the Social: How the Practice Is Responding to Ethical Crisis and Spreading across Sectors.

Theilen, J. T., A. Baur-Ahrens, F. Bieker, R. Ammicht Quinn, M. Hansen and G. González Fuster (2021). Feminist Data Protection: An Introduction. Internet Policy Review **10**(4): 1-26

Titus, A. (2011). Understanding Public Perception: A Network Approach. Background Paper. S. Webb. Newcastle, Australia, Research Institute for Social Inclusion and Wellbeing, University of Newcastle.

Titus, A. (2013). Online Issue Network Analysis. Managing Low Emissions Coal Technologies project risk: The role of public awareness. D. Kuch, A. Titus, J. Ladwig, S. Webb and D. Kennedy. Prepared for the NSW Department of Trade and Investment**:** 37 - 52.

UK, A. (2019). Written Evidence Submitted to the Treasury Committee Inquiry: Consumers’ Access to Financial Services.

Vaithianathan, R., E. Kulick, E. Putnam-Hornstein and D. B. Prado (2019). Section 7: Allegheny Family Screening Tool: Methodology, Version 2. Developing predictive models to support child maltreatment hotline screening decisions: Allegheny County methodology and implementation.

Van Dijck, J. (2014). Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology. Surveillance & Society **12**(2): 197

Van Dijck, J. and T. Poell (2016). Understanding the Promises and Premises of Online Health Platforms. Big Data & Society **3**(1): 2053951716654173

Veale, M. (2019). Draft Ethical Report Concerning Predictive Modelling in the Allegheny Babies and Families Project.

Veale, M. and I. Brass (2019). Administration by Algorithm? Public Management Meets Public Sector Machine Learning. Algorithmic Regulation In K. Yeung and M. Lodge (Eds.). Oxford, UK, Oxford University Press.

Veale, M. and L. Edwards (2018). Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling. Computer Law Security Review **34**(2): 398-404

Veale, M. B., Irina; (2019). Administration by Algorithm? Public Manage-Ment Meets Public Sector Machine Learning. Algorithmic Regulation In K. Yeung; and M. Lodge (Eds.). Oxford, Oxford University Press.

Viljoen, S. (2021). A Relational Theory of Data Governance. Yale LJ **131**: 573

Von Luxburg, U., R. C. Williamson and I. Guyon (2012). Clustering: Science or Art? Proceedings of ICML workshop on unsupervised and transfer learning, JMLR Workshop and Conference Proceedings.

Wachter, S. (2022). The Theory of Artificial Immutability: Protecting Algorithmic Groups under Anti-Discrimination Law. Tulane Law Review **97**(2): 149

Warren, C. (2002). Qualitative Interviewing. Handbook of Interview Research: Context & Method In G. JF; and J. Holstein; (Eds.), Thousand Oaks: Sage Publications.

Webb, S. A. (2001). Some Considerations on the Validity of Evidence-Based Practice in Social Work. British journal of social work **31**(1): 57-79

Welshman, J. (2013). Underclass: A History of the Excluded since 1880, A&C Black.

Woodiwiss, J. (2018). From One Girl to ‘Three Girls’: The Importance of Separating Agency from Blame (and Harm from Wrongfulness) in Narratives of Childhood Sexual Abuse and Exploitation. Pastoral care in Education **36**(2): 154-166

Yanow, D. (2015). Constructing "Race" and "Ethnicity" in America: Category-Making in Public Policy and Administration, Routledge.

Yanow, D. (2015). Making Sense of Policy Practices: Interpretation and Meaning. Handbook of Critical Policy Studies In F. Fischer;, D. Torgerson;, A. Durnová; and M. Orsini; (Eds.). Cheltenham, UK ; Northampton, MA, USA Edward Elgar Publishing**:** 401.

Yeung, K. (2017). ‘Hypernudge’: Big Data as a Mode of Regulation by Design. Information, Communication & Society **20**(1): 118-136

Yeung, K. (2018). Algorithmic Regulation: A Critical Interrogation. Regulation and Governance **12**(4): 505-523

Yeung, K. and M. Lodge (2019). Algorithmic Regulation, Oxford University Press.

Yin, R. K. (2014). Case Study Research Design and Methods. Los Angeles, California, Sage.

Ziewitz, M. (2011). How to Think About an Algorithm? Notes from a Not Quite Random Walk. Discussion paper for Symposium on 'Knowledge Machines between Freedom and Control'.

Zuboff, S. (2015). Big Other: Surveillance Capitalism and the Prospects of an Information Civilization. Journal of Information Technology **30**(1): 75-89

Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. New York, Public Affairs.