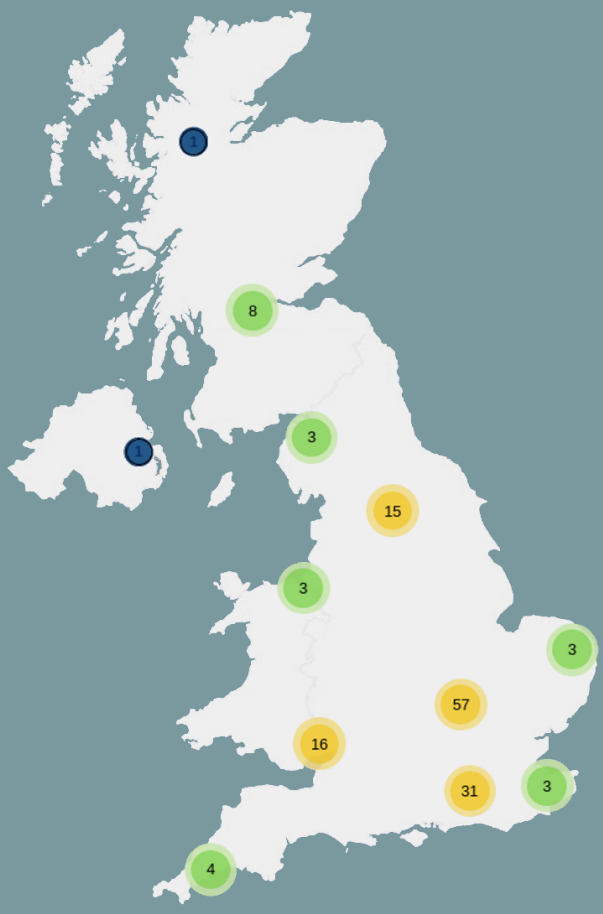
Data Scores as Governance: Investigating uses of citizen scoring in public services

Project Report

Lina Dencik, Arne Hintz, Joanna Redden & Harry Warne



Data Justice Lab

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

OPEN SOCIETY FOUNDATIONS

Data Scores as Governance: Investigating uses of citizen scoring in public services

Project Report

Lina Dencik, Arne Hintz, Joanna Redden & Harry Warne

Data Justice Lab, Cardiff University, UK

December 2018

The project is made possible through a grant from the Open Society Foundations.

Table of Contents

Executive Summary

The project 'Data Scores as Governance' examines the uses of data analytics in public services in the UK. In particular it is concerned with the advent of data-driven scores that combine data from a variety of sources as a way to categorize citizens, allocate services, and predict behaviour. There is an increasing emphasis on data use in UK government and we have seen a proliferation of data systems in both central and local government in recent years. The ability to collect and analyse increasing amounts of data across social life is said to have provided new opportunities to extract new insights or create new forms of value, often through scientific and more objective decision-making. At the same time, the trend of datafication has been met with significant concerns about the characterisation of data as benign and neutral, and attention has been drawn to possible harms of increased monitoring of populations through data, such as infringements upon privacy, potential for discrimination, and an inability to account for complex lived experiences.

The project provides the first comprehensive overview of key developments and outlines concrete examples of how data analytics is implemented and used across different local authorities and partner agencies, what companies and systems are prominent, and understandings and perspectives about these developments amongst stakeholders and civil society. Drawing on multiple methods, including Freedom of Information requests, interviews, and stakeholder workshops, the report details the different ways that data systems are being used in public services. This report is accompanied by the Data Scores Investigation Tool (www.data-scores.org) which is a tool created as part of the Data Scores as Governance project using computational methods to map mentions of data analytics in government.

Our research points to the lack of any systematic information about where and how data systems are being deployed across government. Responses to our Freedom of Information requests illustrate the varied levels of transparency surrounding developments within local government, with no established guidelines in place for interpreting or disclosing information about uses of data analytics to the public. In some instances, extensive detail was provided about the nature of the systems implemented and how it is used, whilst in other, the language used for inquiries about data analytics was not recognised, or information was withheld due to concerns with sensitivity or security.

In collating information about the implementation and uses of data systems in public services, our research indicates that there are no standard practices or common approaches with regards to how data is or should be shared and used by local authorities and partner agencies. Our six case studies looking at areas of fraud, health, child welfare, social services and policing across different parts of the UK paint a varied picture about developments. Whilst data sharing within councils and local authorities is generally a growing trend with the creation of 'data warehouses' and integrated databases - particularly as Councils struggle to respond to austerity measures, rising need and substantial cuts in public services - the use of this data differs significantly amongst different authorities. In the case of Camden's Resident Index, data is combined and analysed to provide a 'single view of the citizen' and is used to detect fraud; in the case of Kent's Integrated Dataset, data analytics is used for public health analysis;

in the case of Bristol's Integrated Analytical Hub, data systems are developed with the view to predict the risk of child exploitation; Hackney County Council is partnering with companies Xantura and EY to use predictive analytics to identify children and families in need of intervention and extra support; in Manchester a Research and Intelligence Database is used for sharing information amongst case workers and other professionals and to carry out network analyses; and in the case of Avon & Somerset Police use of Qlik Sense, data is collected and used to map crime trends and assess the risk of offenders. Our research shows a broad range of data applications. Some of the practices here, such as population level analytics and network analysis, do not involve the production of 'scores'. In other cases scoring can take several forms: in some instances it is predominantly a matching score created for the purposes of identity verification, and in others, it can be an indication that a 'risk threshold' for an individual has been passed and trigger an alert, or it can relate to population level risk assessments.

Our research also illustrates that authorities differ in their use of either in-house or externally developed data systems, with the level of outsourcing being context-dependent. However, we found that a few companies are prominent within the general advancement of data analytics in public services. Desk research and the scoping workshop highlight the prominence of the consumer credit reporting agency Experian and their geodemographic segmentation tool Mosaic in the public sector as a way to analyse populations. We also found that companies such as Xantura, Callcredit and Capita routinely provide data sharing and analytics services to public sector clients across the UK, including profiling systems, identity verifications, and risk assessments.

Whilst stakeholder groups from civil society see benefits with the use of data analytics to provide better and more efficient public services, our research also indicates significant concerns with the way in which some of these data systems are currently implemented and used. In particular, we found concerns with the extent of data collection and sharing, particularly around very sensitive data held by councils; the potential for bias and discrimination in decision-making based on historically skewed data-sets and practices; the possibility for targeting, stigma and stereotyping of particular groups with the labelling of 'risk'; lack of transparency, public knowledge, consent, and oversight in how data systems are being implemented and used; and the limits of data protection regulation to address the overall political context of uses of data in public services to advance particular policy agendas.

In mapping and analysing developments of data analytics in public services our research therefore points to the need for a more nuanced debate about how citizen data is being used and shared. Key is ensuring that such a debate engages both local authorities and impacted communities, that ethical reflections are supported by robust regulation, and that efforts are geared towards identifying avenues of accountability, citizen participation, and possibilities for refusal or non-data solutions when appropriate.

Introduction

The collection and processing of large quantities of data is an increasingly integral part of governance. Whilst the integration of data analytics in government practice is taking many forms, the use of scoring systems and dashboards is a particular emerging practice with significant implications for state-citizen relations. Data scores that combine data from a variety of both online and offline activities are part of a larger trend to use data to identify and categorise citizens, allocate services, and predict future behaviour. Yet little is known about these practices, particularly at the local government level where public services are predominantly provided.

The project 'Data Scores as Governance' examines the uses of data analytics in public services in the UK. It has been carried out by the Data Justice Lab, a research unit situated in the School of Journalism, Media and Culture at Cardiff University, UK. The Data Justice Lab is dedicated to the study and practice of datafication from a social justice perspective, highlighting the politics of data processes from a range of different angles. With this project, we aim to provide the first comprehensive overview of developments pertaining to the implementation and uses of data systems for public services situated within a broader discussion about the challenges of the 'datafied society'. By combining methods, and outlining illustrative examples through in-depth case studies, the project serves to advance scholarly and public debate about these developments, and to point to areas of potential intervention to address citizen needs and concerns.

The report consists of seven main sections based on our research and is complemented by an interactive online tool (www.data-scores.org) to advance further research and investigation into uses of data systems in central and local government. We start by outlining the broader context and background for our research, combining debates prominent within the field of Critical Data Studies with literature on uses of data in public administration, both in the UK and internationally. We then go on to outline our methodology which consists of six aspects: i) desk research, ii) freedom of information requests, iii) stakeholder workshops, iv) practitioner interviews, v) civil society interviews, vi) and computational methods.

The presentation of our findings starts with a qualitative analysis of the Freedom of Information requests and highlights the nature of responses we received and challenges with using this as a method for research. We then go on to outline our case studies, which include six different local authorities and partner agencies across child welfare, health, fraud, and policing, as well as an analysis of the widely used geo-demographic software Mosaic (developed by Experian), and three further company profiles of prominent actors in the supply of data systems for the public sector. Following this, we go on to outline the findings from our interviews with civil society actors, highlighting key attitudes and concerns amongst those working with service-users and impacted communities, as well as key themes from our stakeholder workshops. We end with a discussion of our research, and consider policy implications and possibilities for intervention and citizen participation in the deployment of data systems within the public sector.

Section I:
Background and Context

Background and Context

The use of data analytics in public services is part of a broader development of datafication. As an increasing range of social activity and human behaviour is transformed into data points, it can be tracked and analysed by both commercial and public actors. For governments and public authorities, this holds promise for a better understanding of people's (and broader social) needs, more efficient service allocation, and improved responses to a range of social problems. At the same time, it has also led to concerns regarding the privacy implications of data collection, discriminatory effects of data analysis, democratic accountability, and the wider consequences of the transformation of social life into data.

In this section of the report, we situate our research within these debates on datafication and the use of data analytics within the public sector. The section reviews relevant academic literature and governmental, civil society and media reports that address the opportunities and challenges of the datafication of public services. This will serve as a background for the discussion of our own research results.

Datafication: Opportunities and Concerns

The emerging capacities in analysing 'big data' have led to new opportunities 'to extract new insights or create new forms of value' (Mayer-Schönberger & Cukier 2013, p. 8). Data analysis has enabled the private sector to enhance productivity and has created a new economic sector based on the processing of data about people. This has been hailed as a 'new industrial revolution' (Hellerstein, 2008), with data described as 'the new oil' (The Economist, 2017). Yet data analytics have affected decision-making in a wider range of sectors. Algorithms – automated instructions to process data and produce outputs – may allow for understanding previous occurrences and predicting future behaviour, which may offer opportunities for both private and public governance (Gillespie, 2014). 'If data is the new oil', notes the New Economics Foundation, 'then algorithms are the new refineries' (McCann et al, 2018: 19).

Data analytics promise a scientific and fact-based method for tackling uncertainty. Risks are rendered perceptible through algorithmic calculation and can improve 'proactive' forms of governance (Amoore & Piotukh, 2016). This has been acknowledged in the context of security concerns, but also as an opportunity to enhance the delivery of public services and devise better responses to social problems. As technologically-generated and value-neutral information, data may reduce subjective judgement and thus offer a more rational, impartial, reliable and legitimate way of decision-making (Mayer-Schönberger & Cukier, 2013).

However, these assumptions have been critically interrogated by scholars, particularly in the emerging academic field of critical data studies. A key concern has addressed the characterisation of data as benign, neutral and objective that reflects 'the world as it is' (Kitchin & Lauriault 2015, p. 3). Rather, as critics note, data is always constructed based on the goals, interests and cultures of institutions and individuals (incl. case workers, department heads, and the developers of algorithms), and the perceived objectivity and neutrality of data have been criticised as 'carefully crafted fictions' (Kitchin, 2014). This also means that the representation of 'reality' by data and, more specifically, the relationship between people and the data that are collected about them is not self-evident (van Dijck, 2014). Data analytics

may provide a reduced lens on society (Berry, 2011) and shape the reality it measures by focusing on specific objects, methods of knowing, and understandings of social life (boyd & Crawford, 2012; Cheney-Lippold, 2017). Rather than representing society, data may construct it – as Kitchin (2017, p. 25) notes, data 'are engines not cameras.'

Further, critics have highlighted the risks and implications of increased monitoring and surveillance of populations through data (Van Dijck, 2014; Lyon, 2015) and have analysed a wider range of harms, such as discrimination, that may be caused by the use of big data (boyd & Crawford, 2012; Gangadharan et al., 2015; Redden & Brand, 2018). They have raised concerns regarding the 'operative logic of preemption' (Massumi, 2015) inherent in data-based governance that challenges practices and understandings of the democratic process (Andrejevic, 2017) and focuses on managing the consequences, rather than seeking to understand underlying causes, of social ills (Lyon, 2015). The predominant security discourse of the past two decades has been identified as a source of the 'risk management' focus of many data analytics systems (Coaffee & Murakami Wood, 2006; Aradou & Blanke, 2015).

Uses of 'Big Data' in Public Administration

Government departments and state agencies in many countries now apply data analytics to inform policy and decision-making in a variety of areas. In one of the most prominent studies, Eubanks (2018) has outlined the rise of a 'regime of data analytics' in public services, detailing uses of, for example, automated welfare eligibility systems and the use of predictive risk models in child protective services. In the UK, the increased use of data within government has included the early application of Customer Relationship Management software and similar systems within local government settings (King, 2007), the promotion of "open data" schemes across the UK (HM Government, 2012), and, more generally, the government's push to make public services 'digital by default' (Cabinet Office, 2012). More recently, artificial intelligence and predictive analytics have been used by the Government Digital Services (GDS), for example, to predict future pension scheme behaviour and automate a variety of processes and services (HM Government, 2017). The House of Lords Select Committee on Artificial Intelligence has highlighted 'the Government's leadership in the development and deployment of artificial intelligence' and advised 'greater use of AI in the public sector' (Select Committee on AI, 2018, p. 69).

Several studies have explored the advantages and opportunities of the use of data processing and predictive analytics in the public sector. Nesta's 'Local Datavores' series, for example, has analysed how local authorities can use data and analytics to improve the lives of people and communities. The report 'Wise Council' provides numerous examples and presents lessons that may be learned (Symons, 2016b), while 'Datavores of Local Government' offers advice to councils on how to use the data at their disposal most effectively (Symons, 2016a). They show how data can inform decision-making, problem-solving, and enable changes to practices on the ground. The development of an internal data infrastructure that enables the linking and combining of data from multiple sources, as well as protocols for sharing data, is regarded as core factors for effective data use. Malomo & Sena (2016) add the need for a general legal framework to facilitate data sharing between local authorities, the upgrading of staff skills, and internal institutional support structures. Existing efforts, such as Kent County Council's

approach of linking databases and performing analytics upon their datasets, are highlighted as positive examples (e.g., Symons, 2016b; Malomo & Sena, 2016). The Open Data Institute (ODI) has complemented this perspective with a focus on the social and economic values of open data, which can be enhanced through data analytics (Open Data Institute, 2018).

In the context of challenges to public finances and continued budget cuts, data analytics have also been promoted as a way to reduce costs whilst maintaining or even improving the level of service. Nesta's 'Wise Council' report highlights 'the better use of data and analytics' as 'essential ingredients' to 'address public sector funding and service pressures' (Symons, 2016b, p. 10). The ODI notes that '[m]any public services in the UK are expected to deliver efficiency savings along with improved outcomes for citizens' (Open Data Institute, 2018: 7), while Malomo & Sena (2016, p. 3) point to the funding-related challenge for public authorities 'of having to re-organize their services so that costs can be reduced while simultaneously managing the demand for their services.' The efficiency gains of data analytics are thus particularly attractive in a context of austerity.

The concerns with data analytics, however, as noted above, have been extended to their application in the public sector. Increased collection, analysis and sharing of personal data may lead to 'structural surveillance' (Vagle, 2016) of the population and to a form of social ordering (Lyon, 2015) which may entrench social and economic inequality. Data systems have produced, in some cases, discriminatory outcomes, for example by using skewed data sets that incorporate historical bias into the decision-making (Crawford, 2013), or by repurposing, combining and re-aggregating, and thus decontextualizing, data (boyd & Crawford, 2012). Technical errors have led to the denial of social services to many people who would be entitled to them (Eubanks, 2018). The black boxed nature of big data processes, i.e. the inherent limitations for understanding, investigating and challenging them, poses a significant problem for populations that are assessed by them and whose services are affected by them, particularly in democratic societies (Pasquale, 2015).

Studies about the use of data analytics in public services have identified many cases in which care, social benefits, and other entitlements were drastically reduced after the introduction of data analytics, without a chance for affected individuals to understand or challenge these changes (McCann et al, 2018; Eubanks, 2018). In the wake of scandals, such as Cambridge Analytica's use of Facebook data, and revelations such as those on data and transparency issues of the NHS England's care.data scheme, increased interest has emerged in the potential problems of data analytics (Knapton, 2016; Big Brother Watch, 2014; The Guardian, 2018). Nesta and others have sought to respond 'to people's concerns about a loss of control over their personal information' by developing models for personal data control and have advanced debate on the data commons (Bass et al., 2018). The organisation Involve, in a report on 'Data for Public Benefit: Balancing the risks and benefits of data sharing' (Scott, 2018) has developed a taxonomy of the various implications of data sharing for different stakeholder groups (individuals, communities, service providers). They highlight risks such as 'stigma and discrimination' and 'impacts on communities from the selective use of data' and recommend purposeful, proportionate, and responsible data use, as well as public involvement in conversations around data to enable an 'informed and meaningful dialogue

with service providers regarding their aspirations for how public services should be provided and their concerns about how data about them should be used' (Ibid.: 43). Failure to do so would, in their view, 'undermine the reputation of public service agencies, hamper their ability to resolve disputes, and ultimately constrain their ability to use data in modern, beneficial and potentially transformative ways (Ibid.: 44). The Royal Society and The British Academy, in a joint report, have urged further consideration of the challenges of algorithmic governance. They claim 'high-level principles' are needed to shape all forms of data governance, underpinned by the promotion of 'human flourishing' (Royal Society & The British Academy, 2017). The proposed principles include the protection of individual and collective rights; transparency, accountability and inclusivity of data management, and the enhancement of democratic governance.

The prevalence of public-private partnerships is a further characteristic and, for some, concern regarding the use of data analytics in public services. Despite the highly influential role government funding often plays in the creation of new technologies, the private sector has often been at the forefront of the application of data-focused solutions, and many government agencies use services developed and supplied by commercial providers (Mazzucato, 2018). This has exacerbated concerns regarding the transparency and accountability of data systems as the algorithms at the centre of data analytics have often been regarded as trade secrets and have thus been exempt from public scrutiny. It has, more broadly, underlined critiques of data analytics with regards to democratic control if automated systems of state-corporate decision-making affect the lives of populations, with reduced possibilities for citizens of remedy and participation (Zuboff, 2015). While the use of tools and services, such as Experian's geodemographic segmentation system, Mosaic, is widespread within the public sector, this is rarely discussed in literature promoting the use of data analytics in government, which suggests a normalisation of private sector involvement and a lack of awareness of related challenges and implications.

Data Scores

The use of data analytics in public services typically involves practices of categorizing and segmenting, and sometimes rating and ranking, populations according to a variety of datasets, with the goal of allocating services accordingly and identifying specific 'risks' and behaviours. Data scores that combine data from different sources towards a numerical index are emerging as a prime means for such categorizations.

In the commercial realm, the practice of scoring is well-known from the financial sector and the practise of producing financial credit scores which assess an individual's creditworthiness. The history of the individual's financial transactions in the past is thereby used to predict their likely financial responsibility in the future. A wider range of consumer scores are now being applied across different economic sectors (Dixon & Gellman, 2014). However, the financial sector has also started to experiment with expanding the sources of data to introduce more socially oriented judgements into financial decision-making processes. These may include, for example, an analysis of people's mobile phone use, or the creditworthiness of their social media friends. People's social activities are thus increasingly incorporated into particular commercial assessments, which points to a growing integration of social and transactional

datasets (Fisher, 2018; McCann et al, 2018, p. 10). This practice builds on established experiences in the marketing industry and, more recently, the platform economy, where consumption patterns are predicted based on a variety of social, cultural, health and other data.

Yet the use of data scores has reached governmental and public services, too. In education in the United States, data scores support personalized learning and individualized instruction of students (Crooks, 2017). Data collected on teachers is used to algorithmically score their performance (O'Neill, 2016). In criminal justice systems, risk assessment tools are used to produce 'risk scores' on defendants to estimate their likelihood of re-offending and thus determine sentencing (Angwin et al., 2016). In border control, data-driven profiling based on a cross-set of aggregated data is increasingly used for 'vetting' the 'threat' of migrants and refugees to society, producing what some refer to as a 'terrorist credit score' (Crawford, 2016). Further examples of 'risk' scoring have emerged in the health and family sectors (Warrell, 2015; Tucker, 2016). Recent debates on the use of data analytics in the UK have addressed, among others, the Durham Constabulary's Harm Assessment Risk Tool (HART) and its usage of data provided by the consumer credit reporting agency, Experian (Big Brother Watch, 2018), and data-based decision-making in the field of child welfare (McIntyre & Pegg, 2018a; 2018b; 2018c).

The most comprehensive scoring system is currently developed in China. The 'social credit score' aims at integrating the rating of citizens' financial creditworthiness with a wide range of social and consumer behaviour to assess people's overall trustworthiness and allow, or deny, services accordingly. Government plans were outlined in 2014, with the planned scoring systems to come into effect in 2020. Pilot schemes in specific provinces have involved the local government awarding people points for what is regarded as good behaviour (such as community engagement and donations to charities) and deducting points for negative behaviour, such as traffic offences or spreading online 'rumours'. Citizens with high scores receive privileges, such as fast-track promotion at work or access to good schools and housing, while those with low scores are restricted from, for example, certain forms of travel. The pilot scheme in Suining assigns citizens a grade from A to D, whereas the private system Zhima Credit assigns a numerical score between 350 and 950 (Hvistendahl, 2017).

The social credit score system has been criticised as a 'digital totalitarian state' (The Economist, 2016) and a 'tool for social control' (Chin and Wong, 2016) due to its comprehensive and punitive nature and the lack of limits of data collection and use. While acknowledging that both the context and the goals for scoring are very different in Western countries, observers have noted that scoring is, in principle, becoming more widespread in the West, too – 'it's just distributed between a range of competing services and often shrouded in corporate secrecy' (Silverman, 2015). Some aspects of the system relate to general characteristics of scoring as noted above. The score categorises the citizenry and assigns distinct services across different bands of the score (e.g., different privileges for those over 700 compared to those in the 600s and those in the 500s). It combines different types of data from, e.g., online consumption; use of services; legal, financial and educational records; social media activity; etc. It is based on a public-private partnership, with the government

enlisting major tech companies such as Baidu, Alibaba and Tencent to develop relevant databases, provide user data and incorporate the system into their services. According to Lv & Luo (2018, p. 3890), these internet giant's access to technology and data 'makes them indispensable […] in the building of the social credit system'. The social credit score thus serves as a unique example of scoring that is, in many ways, distinct from the cases discussed in this report, but it demonstrates possible implications of the algorithmic mediation of daily life and therefore offers one of the useful starting-points for investigating the use of data analytics in the public sector of other countries (Fullerton, 2018; Jefferson, 2018).

The Need for Understanding Data Scores

As we have briefly summarised here, there is an emerging set of experiences regarding the use of data analytics in public services. There is also a growing media interest, fuelled, not least, by prominent cases such as the Chinese social credit score. And while the use of data promises enhancements in public service provision, there are significant concerns regarding the datafication of social life and implications for citizens. Yet there is a lack of systematic reviews of data analytics in the public sector and, specifically, of scoring systems. Some research has emerged regarding the use of algorithms in public services in the US (Eubanks, 2018; O'Neil, 2016; Diakopoulos, 2014; Angwin et al, 2016). However, very little is known, so far, about the role of data scores and algorithmic decision-making in the UK public sector (with a few notable exceptions, e.g. Hall & McCann, 2018; McCann et al, 2018).

The lack of public knowledge and public debate as well as systematic public sector engagement with the issue has been noted, among others, by the House of Commons Science and Technology Committee in a report on algorithms in decision-making. The Committee recommends that 'The Government should ... produce, maintain and publish a list of where algorithms with significant impacts are being used within Central Government, along with projects underway or planned for public service algorithms, to aid not just private sector involvement but also transparency. The Government should identify a ministerial champion to provide government-wide oversight of such algorithms, where they are used by the public sector, and to co-ordinate departments' approaches to the development and deployment of algorithms and partnerships with the private sector' (Science and Technology Committee, 2018, p. 3). Further, the need to expand the public's, journalists', and academia's ability to scrutinise the use of data scoring systems and other forms of data analytics within the public sector has been highlighted by foundations, such as the Omidyar Network and Upturn, that claim 'a clear agenda for public scrutiny has yet to emerge' (Omidyar Network & Upturn, 2018, p. 30).

This project therefore seeks to advance public, academic, and public sector understanding of data analytics by providing new evidence, compiling case studies and experiences, and addressing both opportunities and challenges. This report will offer accounts of where, how and to what ends data analytics have been applied in public services in the UK, discuss their effectiveness for public services, and explore how data driven decision making has the capacity to influence lives and service outcomes. More research in this area is required if we are to foster a truly informed public discourse, but this project can, we hope, offer a useful building block.

Section II:
Methodology

Methodology

At present central and local governments across the United Kingdom do not provide lists of how algorithmically driven systems are influencing services and decision making. This project begins from the position that such a record is necessary. As noted in the previous section, it is a position also taken by the House of Commons Science and Technology Committee, which argued in 2018 that:

> The Government should ... produce, maintain and publish a list of where algorithms with significant impacts are being used within Central Government, along with projects underway or planned for public service algorithms, to aid not just private sector involvement but also transparency. (Science and Technology Committee, 2018: 3)

Our project has attempted to address this lack by: a) beginning to build such a record, and b) highlighting the difficulties of such a task, therefore indicating the need for further action.

We took inspiration from a number of projects, particularly the Algorithm Tips[1] project and Brauneis & Goodman's (2018) work on transparency around governmental deployment of big data analytics. In line with this work, we have made use of a multi-methods approach. This has involved:

1) Desk research
2) Freedom of Information requests
3) Workshops with stakeholders
4) Practitioner interviews
5) Civil society interviews
6) Computational methods (data scraping)

As our goal was both building a record of systems and investigating how they work, we deployed a broad definition of data systems which included the uses of large integrated datasets as well as predictive analytics and scoring systems. We have attempted both a general mapping of systems and practices as well as more detailed case study investigations. The general mapping of systems culminated in the Data Scores Investigation Tool (www.data-scores.org) and the case study investigations include:

a) Bristol's Integrated Analytical Hub
b) Kent's Integrated Dataset
c) Camden's Resident Index
d) Hackney's Early Help Profiling System

[1] http://algorithmtips.org/

e) Manchester's Research & Intelligence Database

f) Avon & Somerset Police Qlik Sense

In addition, we carried out a number of investigations into prevalent companies involved in public sector data analytics. These include:

i. Experian (Mosaic)

ii. Xantura

iii. Callcredit

iv. Capita

Desk research

We began by focusing on the uses of data systems by local government because much social and health service administration and use happens at the local level. Early desk research included sampling media coverage of data analytics and algorithms using Lexis Nexis, and carrying out general search queries of local and central government websites. This process identified a number of systems in operation and being developed. This initial phase of the research informed the search terms for the computational methods, the articulation of the Freedom of Information requests, and the sampling of case studies for both local authorities and companies along with detailed background information.

Freedom of Information requests

Freedom of Information (FOI) requests were sent to each case study relating to a system we had identified through desk research. A copy of a request can be found in the Appendix 1. The targeted requests were partially successful, with some responses providing a great deal of detailed information. In order to develop a more comprehensive account of the range of algorithmically driven data systems in use across local authorities (given also our resource and time constraints), we sent out general FOI requests to local authorities and agencies across the UK, totalling 423 requests. For this we used WhatDoTheyKnow Pro, a premium offering from WhatDoTheyKnow.[2] This is an online service which simplifies the process of submitting UK Freedom of Information requests run by the non-profit organisation mySociety.

The use of FOIs to investigate the integration of changing data systems is problematic and resource intensive for all parties. However, in the absence of a public list, the Freedom of Information Act provides an opportunity for systematic inquiry. We have sought to provide a more qualitative analysis of these requests in this report and the responses to requests have been added to the publicly accessible Data Scores Investigation Tool discussed below.

2 www.whatdotheyknow.com

Workshops

As part of the project, we organised a number of workshops with different stakeholder groups:

1) Scoping workshop (20 April 2018)
2) Journalist workshop (21 September 2018)
3) UN investigation workshop (6 November 2018)
4) Project findings workshop (19 November 2018)

The Scoping workshop included participants from councils, local authorities, emergency services, think tanks, civil society organisations and academia with the view to scope developments and debates relating to uses of data analytics in the UK public sector. The workshop made use of Chatham House rules and was divided into three consecutive sessions which were designed to broadly reflect the concerns and perspectives of government, the research community, and civil society: 1) Data analytics in the public sector: Experiences, opportunities and challenges; 2) Exploring data uses: Research and reviews; 3) Civil society perspectives: Challenges and concerns. This workshop provided key insights into developments, possibilities and concerns surrounding the use of data analytics in public services and practices relating to citizen scoring which informed subsequent research.

The Journalist workshop brought together national and regional journalists within the UK, together with data journalism educators and civil society actors. The aim of the workshop was to test the Data Scores Investigation Tool amongst key users and gather feedback to improve design and features, as well as gather insights on how to advance an 'algorithms beat' in journalism education and practice. The workshop consisted of a tool demonstration, interactive user-tests, and interventions from journalism educators.

The UN investigation workshop was organised in connection with the visit from the UN Special Rapporteur for extreme poverty and human rights to the UK to investigate the effects of austerity. As part of this investigation, the workshop brought together frontline staff and civil society organisations to share experiences and views on the use of digital technologies in welfare provision and possibilities for enhancing citizen participation in the deployment of data systems in the public sector.

The project findings workshop brought together public sector workers, civil society groups, journalists and scholars to discuss preliminary findings from the project at an event held in Westminster. Following a presentation of the findings from the project research team, two sessions followed organised around 1) stakeholder perspectives and 2) policy implications. The workshop invited general reflections on the wider implications of the project.

Practitioner interviews

Our case study investigations included practitioner interviews with actors working with the implementation and uses of data systems in local authorities and partner agencies. For each case study, we sought to interview people involved with the development, management, or

user side to include a range of perspectives. Interviewees were sampled through desk research identifying actors in key roles relating to each case study. In total, we interviewed 17 practitioners working across our six case studies, either in person or on the phone, lasting on average around an hour. The interviews explored questions about the details of the system, the terms of its implementation and uses of the system for each case study. A sample interview guide is included in Appendix 3.

Civil society interviews

In addition to practitioner interviews, we conducted 10 interviews with people from a range of civil society groups that were sampled as public service stakeholders and having familiarity with impacted communities. These included a range of orientations pertaining to digital rights, welfare rights, and citizen participation. Interviews were carried out in person, through online video or on the phone, lasting on average 30-45 minutes, and explored questions about the knowledge of developments in data analytics for public services, opportunities and concerns, and avenues for addressing and mitigating risks and harms. See Table 1 for a list of the organisations interviewed and see Appendix 4 for a sample interview guide.

Table 1 *Civil society organisations interviewed*

Organisation	**Orientation**
Big Brother Watch	Civil liberties
British Association of Social Workers (BASW)	Professional association
Citizen's Advice Bureau	Advice & advocacy
Defend Council Housing	Housing activism
Disabled People Against the Cuts (DPAC)	Disability activism
Involve	Public engagement
Liberty	Human rights
Netpol	Police watchdog
Open Rights Group (ORG)	Digital rights
Welfare rights activist	Troubled Families

Computational methods

Drawing from the methodology applied in the Algorithm Tips project[3], we used computational methods for our research with the view to map developments across the UK. To do this, we used search engines to scrape UK government sites and media sites based on a list of keywords. We adapted the keywords used by Algorithm Tips, adding words we thought more likely to return productive results in a UK context. We also included the names of software and companies we had encountered in our desk research. The data also includes material gained through our freedom of information requests. This resulted in the Data Scores Investigation Tool. A prototype of the tool was tested with journalists and civil society users, and feedback was used to improve features and design.

The tool is available at www.data-scores.org. For a code repository which contains the resources required to replicate the tool, visit:
https://github.com/critocrito/data-scores-in-the-uk

[3] http://algorithmtips.org/methodology/

Section III:

Freedom of Information Analysis

Freedom of Information analysis

Given the lack of publicly available information on the use of data analytics and algorithmic systems in the UK public sector, we chose to use the Freedom of Information Act 2000 and Freedom of Information (Scotland) Act 2002 to request information from Local Authorities. Our use of freedom of information requests as a means to investigate government uses of algorithmic systems is in line with previous research in the United States. Our approach is informed in particular by work by Brauneis & Goodman (2018) and Fink (2017).

This section details our approach and findings. Our freedom of information requests generated important information but also revealed the limits of using the FOI Act to research data analytics in the public sector.

The process

We began by submitting targeted FOI requests to learn more about the Local Authorities and systems we had flagged as potential case studies (not all of which were pursued in the end as part of our case study research). As our aim was to map, as much as possible, data systems across government we also submitted general FOI requests to local authorities across the UK. We submitted 20 targeted requests and 403 general requests bringing our total to 423.[4] We used WhatDoTheyKnow's "Pro" features[5] to submit these additional requests, whereas our previous requests had been submitted from a Cardiff University email account.

Appendix 2 contains our general, exploratory request. The request contains a few passages which attempted to preempt various issues we had faced with our initial round of targeted requests. In these requests we included links to documents from the Information Commissioner's Office (ICO)[6] and the Department for Communities and Local Government[7] to justify using the name of the Lab rather than an individual (required because multiple researchers could be handling the request) and that Local Authorities should not invoke commercial sensitivity to deny the release of a contract with a private contractor (i.e. the developer of a data system). Our experience was that these preemptions were often ignored or only recognised once we referred to them in a follow up email. This was particularly the case with our signed name, with many Local Authorities requesting a "real name" before they could process our request but who recognized our ability to use an organization name once we referenced the information commissioner guidance.

Types of response

Fig. 1 *A chart of the types of response to our Freedom of Information requests*

[4] A list of Local Authorities was parsed using code written in Python from the .csv file available at this link: https://data.gov.uk/dataset/local-authority-services [accessed 26/01/2018]

[5] https://www.whatdotheyknow.com/pro

[6] https://ico.org.uk/media/for-organisations/documents/1164/recognising-a-request-made-under-the-foia.pdf -- The relevant sections are on pages 9-10, sections 38-39.

[7] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/408386/150227_PUBLICATION_Final_LGTC_2015.pdf -- The relevant sections are on page 9, section 20.

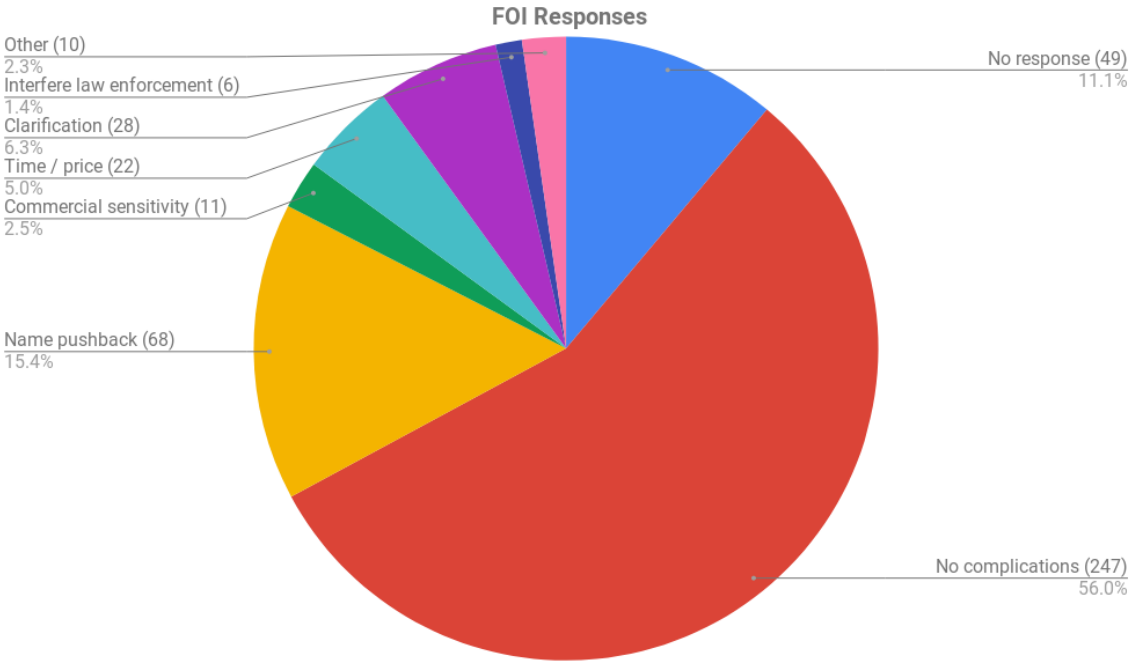


Fig. 1 shows the types of initial responses we received to our FOI requests. We recorded responses according to category, as one response to our request could contain several complications. For example, a response could both request clarification in one area and also insist that we provide the name of an individual as requester. As Fig. 1 indicates just over 60 percent of our total number of 403 requests were processed without complication. The responses to the other 'complicating' requests often meant that we needed to provide a follow-up email before the request could proceed. Some responses relate to information being partially withheld (for example, a council providing us with information but withholding a single document on commercial sensitivity grounds). This chart highlights the objections which slowed down and limited this exercise.

"No response" refers to requests where we did not hear anything from a Local Authority beyond what appeared to be an automated response.[8]

"Interfere law enforcement" refers to where part of the request was refused on the grounds that the release of the information could interfere with the policing of unlawful activity. This was often invoked when it was said the release of information could aid individuals in committing fraud relating to the system we were requesting information about. These instances often invoked Section 31 of the Freedom of Information Act 2000.[9]

"Clarification" refers to instances where an authority asked us for clarification on our request. This usually involved asking for a more robust definition of what we meant by data analytics or another term used within our request (see Appendix 2). This was more prevalent in our general, exploratory request since these requests did not refer to a specific systems. This was

[8] Which is different from a request being acknowledged as valid and the 20 working days time limit for a response beginning.

[9] https://www.legislation.gov.uk/ukpga/2000/36/section/31

expected given the difficulty, often, in communicating about data systems and also because finding out information about existing systems is challenging.

"Time / price" refers to instances where our request was held back by a Section 12 exemption where "the authority estimates that the cost of complying with the request would exceed the appropriate limit".[10] The "appropriate limit" is £450, which works out to 18 staff hours at a cost of £25/hour.[11] Many requests for clarification were an attempt by the Local Authority to avoid invoking this exemption. When Section 12 was invoked the Local Authority often noted how many departments they would have to consult and how many documents they would have to analyse to adequately answer our questions. Some authorities recognised this but still gave us some information that was more easily accessible to them.

"Commercial sensitivity" refers to responses indicating that the release of information would be against the commercial interests of a private company. These often invoked Section 43 of the Freedom of Information Act 2000[12] or Section 33 of the Freedom of Information Act (Scotland) 2002[13], but this was also invoked without reference to either Act. This objection was listed by one Local Authority on advice they had received from a private contractor.

"Name pushback" refers to when a Local Authority requested we provide a "real name" instead of "Data Justice Lab". This is despite our attempts to pre-empt this issue by including advice from the Information Commissioner's Office - which supported our action - in our original request.

Scottish councils proved particularly reluctant, owing to different advice from the Scottish Information Commissioner. We clarified our responses in these cases to include ICO and Scottish Information Commissioner advice, since we believed the Scottish Commissioner's advice allowed us to sign our requests this way. Some Scottish councils accepted our interpretation and processed our requests and some disagreed with our interpretation of the advice. Towards the end of this process we contacted the Scottish Information Commissioner and they agreed with our position but this proved too late in the project to be used in most of this correspondence.

Even when using a tool like What Do They Know which is meant to make the process easier, a great deal of work is involved in managing these requests. Due to the time and resource constraints of this project, at the time of writing, there are still requests which are awaiting clarifications from us or which need their status updated on WhatDoTheyKnow.[14]

[10] https://www.legislation.gov.uk/ukpga/2000/36/section/12

[11] https://ico.org.uk/media/for-organisations/documents/1199/costs_of_compliance_exceeds_appropriate_limit.pdf (p.4)

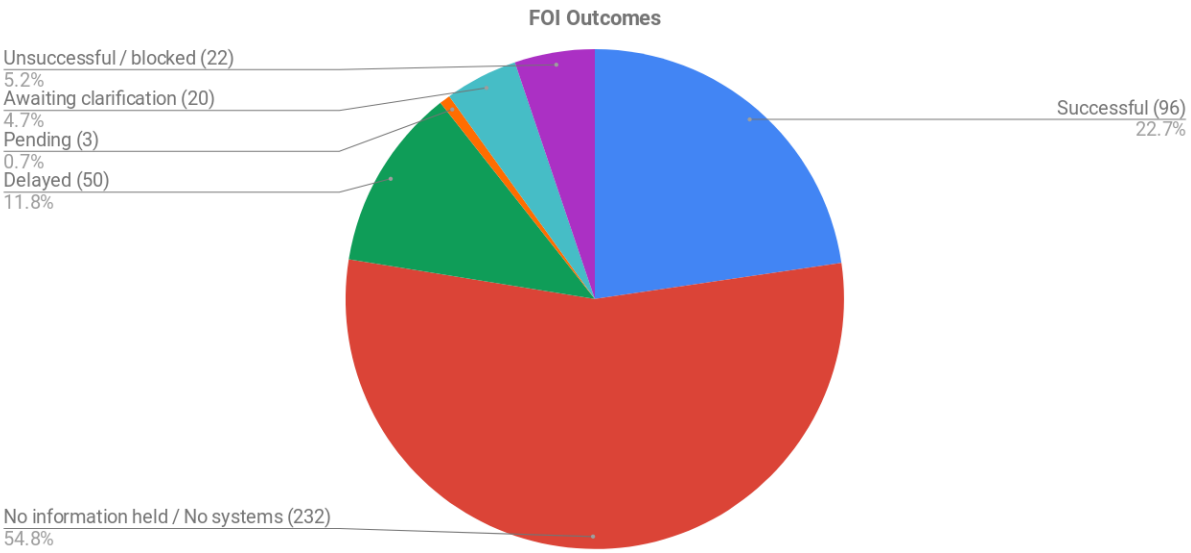
[12] https://www.legislation.gov.uk/ukpga/2000/36/section/43

[13] https://www.legislation.gov.uk/asp/2002/13/section/33

[14] We thank mySociety for all the work they have put into WhatDoTheyKnow which helped make this task much much easier than it otherwise would have been. We would like to give particular thanks to their customer service team who helped us on a number of occasions.

Outcomes

Fig. 2 *A chart of the outcomes of all 423 Freedom of Information requests*



As Fig. 2 shows, most of our 423 FOI requests did not yield results. Included in this chart is data from our 20 targeted case study related requests, sent over the course of a few months from March 2018, as well as data from the 403 exploratory requests we sent on 20th July 2018. Most of our requests received very short responses informing us the Local Authority either held no information on the sorts of systems we were asking for (59) or had no such systems (173). In these cases follow-up is required to determine if in fact there are no data systems in use or if we have not used the right language. Many requests are marked as delayed meaning that, at the time this chart was produced (end of October 2018), the Local Authority had exceeded the 20 working days response time imposed by Section 10 of the Freedom of Information Act 2000.[15] Due to time and resource constraints some requests are still Awaiting Clarification from us. The small number of Pending requests refer to requests where the 20 working days limit has not, at the time the chart was produced, yet been exceeded. Under one quarter of our requests are marked Successful. This proportion was higher with our targeted requests (55%) than our exploratory requests (21%).

Quality of "Successful" responses

We deploy a broad definition of 'Successful' which could refer to anything from a detailed response covering every question within our request, providing accompanying PDFs, contracts, etc., to a single line response only mentioning the name of one piece of common software. We used this broad definition to attempt to capture the diversity that sits beneath a term like "data analytics". At the core of our general, exploratory request were 7 questions (see Appendix 2). Table 1 shows the number of positive responses received (meaning some information or materials were received) and negative responses received (meaning nothing

[15] https://www.legislation.gov.uk/ukpga/2000/36/section/10

was received or question not acknowledged) within the 85 exploratory FOIs[16] marked 'Successful'.

Table 1 *The number of positive and negative responses we received against each question in our Successful exploratory FOI responses*

Question	Positive	Negative
Briefings, reports etc.	12	73
Contracts	10	75
Overviews re: system outputs	9	76
Promotional materials, etc.	9	76
Training manuals, etc.	6	79
Data visualisations	5	80
Impact assessments	3	82

As this table indicates for many responses marked 'Successful' we would often not have every aspect of our request addressed. In a different context this could have been chased up by us or an internal review could have been requested but given the number of requests we were submitting, the extra labour required for this would have exceeded the resources available for this project. Table 2 shows the number of the 85 Successful responses which did not acknowledge each question.

Ninety-six of our requests yielded some information about the use of data analytics. All 'Successful' responses and their accompanying attachments have been incorporated into the database of the research tool we produced as part of this project, which can be found at www.data-scores.org. Table 3 lists the data analytics systems within our Successful FOI responses which were named by more than one Local Authority.

Appendix 5 contains a complete list of the systems or related processes mentioned within our Successful responses, alongside the name of the Local Authority said to use the system. We have also included free text responses where a system has been referred to but not given a name. For the exploratory requests, we have also included a link to the request on WhatDoTheyKnow. The list is alphabetised by Local Authority name. The list may be of interest to anyone wishing to research data analytics systems at the local or national level.

Table 3 *Systems mentioned at least twice within our Successful FOI responses*

Count	System
19	Risk Based Verification (multiple providers)

[16] Most targeted FOIs had their own, unique set of questions and so have been excluded from this part of the analysis.

7	Power BI (Microsoft)
6	Business Objects (SAP)
6	GIS (Geographical Information System) [generic term]
5	Excel (Microsoft)
5	Google Analytics
5	Mosaic (Experian)
5	Tableau
4	National Fraud Initiative (NFI)
3	Capita Revenues and Benefits software
3	CapitaONE
3	Care First
3	IDEA Data Analysis (CaseWare)
3	NHS Health Check
2	ACORN (CACI)
2	Business Intelligence tools [generic term]
2	Crystal Reports
2	Dynamics (Microsoft)
2	Experian Public Sector profiler
2	FACE (Imosphere)
2	Google Tag Manager
2	SSRS

FOI Discussion

Our two biggest takeaways from this experience are:

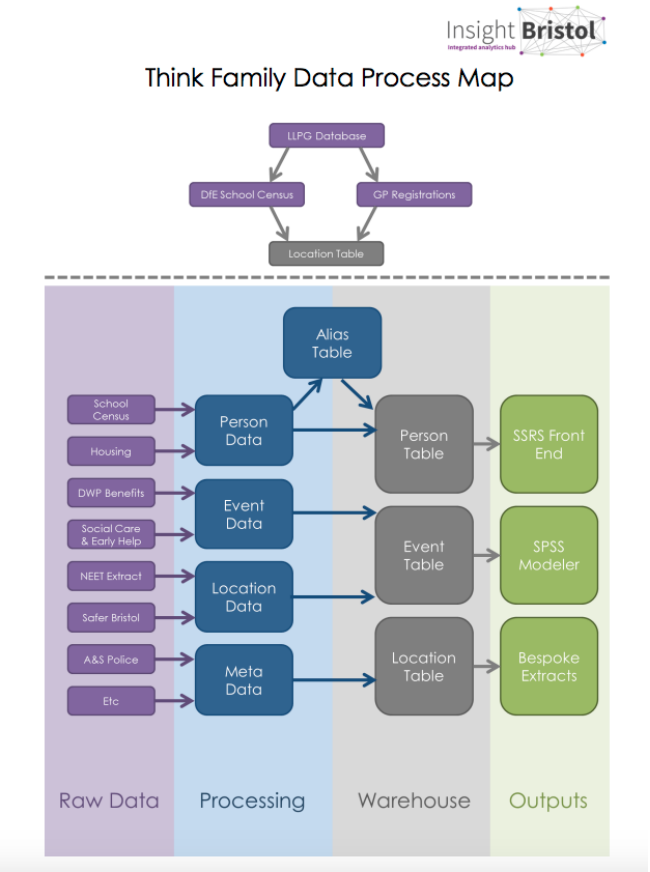
1) The Freedom of Information Acts, as they are currently constituted and implemented, are an imperfect tool for mapping government data systems. However, this method does enable breadth of access and systematicity. We argue that government bodies should compile and maintain a database of data analytics systems used by Local and Central Government. This echos the recommendations of the House of Commons Science and Technology Committee. The Committee recommended Government "produce, maintain and publish a list of where algorithms with significant impacts are being used within Central Government" (Science and Technology Committee, 2018: 3). Our research indicates this should also cover other areas of the public sector.

2) There is need for more follow-up and further research in relation to the use of FOI's and the analysis of the responses to our FOI requests. We hope the information produced assists further research and debate.

Section IV:

Case Studies

Bristol Integrated Analytical Hub



Summary

The Integrated Analytical Hub is an in-house developed system that was established out of Bristol City Council's Think Family approach to the national Troubled Families programme 'to encourage services to deal with families as a whole, rather than responding to each problem, or person, separately'[17]. The Troubled Families programme was launched in 2011 to help families who struggle with factors such as unemployment, crime and poor school attendance. Think Family identified families facing issues, such as parents and children involved in crime or anti-social behaviour; children not attending school regularly; children who need help; adult out of work or at risk of financial exclusion and young people at risk of worklessness; families affected by domestic violence and abuse; parents and children with a range of health problems. Bristol's Think Family programme is now in its second phase. As a result of the learning from the first phase, Bristol developed a Think Family Database that consolidated information from 35 different social issue datasets, about 54,000 families in the local authority area, to understand the strategic and operational needs of the city. The database is up and running and is now able to provide information to predict future need as well as simply responding to presenting issues.[18]

[17] https://www.bristol.gov.uk/policies-plans-strategies/the-troubled-families-scheme

[18] https://www.bristol.gov.uk/documents/20182/34776/Bristol+City+Council+Think+Family+PID+Integrated+Ana

Implementation

The Integrated Analytical Hub is developed in-house by staff employed by Bristol City Council. This was explained as being about having 'complete control over everything' and a concern that 'if you get somebody else with a black box, no one really knows how it works. Assuming you've got to explain it to somebody, you've got no chance.' (data scientist) The decision to develop it in-house also came from a pragmatic concern with using an iterative model based on 'the existing IT' and that 'wouldn't have high level maintenance costs going into the future' (manager).

Initially, the hub was created as a 'data warehouse' around the Troubled Families programme, that combines data of 'all social issues' across the city for children and families to provide a 'holistic understanding' of the family (manager). The decision to create a data warehouse came from a perceived need to 'have a more strategic understanding of the city, the challenges the city faced, the families faced and understand that…[to] make better strategic plans, make better resource decisions, analyse the information better to understand where the risk and vulnerability was in the city, and who was working with those people.' (manager)

Once the data warehouse was created, the team began looking into doing 'more interesting things' such as predictive modeling, particularly in relation to Child Sexual Exploitation (CSE). In a document from Bristol City Council obtained as part of an FOI request, the CSE Model is reasoned as:

> Predictive modeling allows us to make better use of data, to understand the known issues affecting the citizens of Bristol now and in turn how these factors may develop in the future. Understanding future trends can inform resource allocation and decision making on both strategic and operational levels.

As stated by a senior manager, the ability to do predictive analytics relies on having created a data warehouse first that combines data for all children and families around social issues:

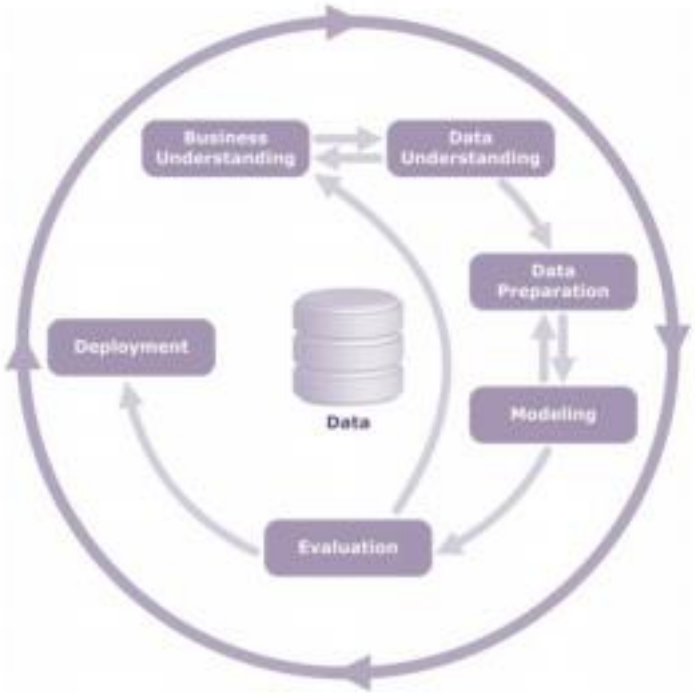
> I thought actually what you could do is start to use predictive analytics around social issues. So could you look at the wider factors that are supporting the reason why a child or young person would demonstrate certain behaviours and understand those characteristics and then start to look at that from that perspective. (manager)

This requires an understanding of characteristics within the cohort who demonstrate the behaviour in relation to the wider population. 'So in order to do that predictive analytics, you needed a data warehouse of all the social issues.' (manager) Moreover, the predictive model is a way to overcome discretion and reliance on professional judgement only: 'it was self-evident and common-sense that there were certain social situations that were precursors to demonstrating future types of behaviour and whilst it was self-evident to the professional…what we didn't have was any of that on any scientific footing.' (manager)

---

lytical+Hub.pdf/3ab01433-e1a7-4e3e-9746-bfcca2516095 (document now removed from website, a redacted version is available in an FOI request)

Model



The cycle shown in the diagram represents steps in developing a predictive model. It highlights the importance of testing and the cyclical nature of the process following feedback from the business (obtained through FOI request). This is informed by what is described as 'co-design'. In the case of the Bristol CSE model, partners from the Barnados project BASE (Barnardo's Against Sexual Exploitation) were consulted while gathering evidence and information on known cases of CSE in Bristol.

The model is based on 35 different social issue data-sets (two of which are not currently collected, see below).

Code	Indicator Name	Source	Legal Gateways
Crime & ASB			
1A	Child / Adult committed offence	ASC	1, 5, 9, 10, 11, 12
1B	Child / Adult committed ASB	BCC & ASC	1, 5, 9, 10, 11, 12
1C	Adult Prisoner 12 Months from release	ASC	1, 5, 7, 9, 10
1D	Adult Subject to license or supervision	ASC	5, 7, 9, 10
1E	Child / Adult serving ABC's, Community Order or Restorative Justice	ASC	1, 5, 7, 9, 10, 11, 12
1F	Child / Adult Professional referral	Lead Professional	5, 7, 9, 10, 11, 12
School Attendance			
2A	Child persistently absent	BCC	1, 4, 5, 8, 9, 10, 11, 12
2B	Child 3 fixed term exclusions	BCC	1, 4, 5, 8, 9, 10, 11, 12
2C	Child permanently excluded	BCC	1, 4, 5, 8, 9, 10, 11, 12
2D	Child Pupil Referral Unit	BCC	1, 4, 5, 8, 9, 10, 11, 12
2E	Child not registered	BCC	*Not currently collected*
2F	Child Professional Referral	Lead Professional	8, 9, 10
Children Needing Help			
3A	Child Active LCS/EHM Episode	BCC	1, 2, 4, 6, 8, 9, 10, 11, 12
3B	Child looked after	BCC	1, 2, 4, 6, 8, 9, 10, 11, 12
3C	Child in need	BCC	1, 2, 4, 6, 8, 9, 10, 11, 12
3D	Child protection plan	BCC	1, 2, 4, 6, 8, 9, 10, 11, 12
3E	Child teenage pregnancy	NHS	1, 2, 4, 6, 8, 9, 10, 11, 12
3F	Child missing	ASC	1, 2, 4, 6, 8, 9, 10, 11, 12
3G	Child / Adult Homeless	BCC	1, 2, 3, 4, 6, 8, 9, 10, 11, 12
3H	Child risk of sexual exploitation	ASC	1, 2, 4, 6, 8, 9, 10, 11, 12
3I	Child Professional Referral	Lead Professional	2, 8, 9, 10
Financial Exclusion & Worklessness			
4A	Adult out of work benefits	BCC & DWP	1, 9, 10, 13
4B	Child risk of NEET	BCC	1, 4, 6, 8, 9, 10, 11, 12
4C	Child NEET	BCC	1, 4, 6, 9, 10, 11, 12
4D	Child / Adult Professional referral	Lead Professional	4, 6, 9, 10
4E	Child / Adult Benefit Cap	BCC	1, 9, 10, 13
Domestic Abuse			
5A	Child / Adult victim of DVA	ASC	1, 4, 7, 9, 10, 11, 12
5B	Child / Adult perpetrator of DVA	ASC	1, 4, 7, 9, 10, 11, 12
5C	Child / Adult domestic incident police callout	ASC	4, 7, 9, 10, 11, 12
5D	Child / Adult Professional referral	Lead Professional	7, 9, 10
Health Problems			
6A	Adult mental health problem	BCC	1, 7, 9, 10, 11, 12
6B	Child / Adult drug or alcohol incident	BCC	1, 4, 7, 9, 10, 11, 12
6C	Adult drugs / alcohol support	BCC	1, 7, 9, 10, 11, 12
6D	Adult receiving Universal Partnership Plus		*Not currently collected*
6E	Child / Adult Professional referral	Lead Professional	1, 9, 10

In interviews it was noted that the Council also buys some data in from external companies such as Experian, for a Joint Strategic Needs Assessment (JSNA), which is a needs analysis based on ward based area data 'to look at population growth and various things.' (manager) The system is running on a combination of software, including SSRS for front line staff accessing the system, SPSS for predictive modeling and Qlik Sense for MI and analytical functions. The model produces an automated risk score from 0 to 100 for every young person in the database. This score is based on statistical probability of similarity in relation to characteristics from a training cohort. This training cohort consists of confirmed victims of

exploitation over the last 2-3 years that have been profiled in order to assess how statistically similar young people in the database are to this training cohort. This process was described in an interview as being a form of 'mirroring': 'it's not predicting you will be sexually exploited or whatever, it's saying you are demonstrating exactly the same characteristics and behaviours as someone who was sexually exploited.' (manager)

The model is now integrating more real-time data into the score produced, by connecting to universal services and receiving daily data on factors such as school attendance:

> at midday we'll be able to see if somebody attended school this morning…it's a lot more useful because it's got some value to it whereas social workers were, why are you telling me someone didn't attend school in December? That's not interesting for me because I need to know what they're doing now. (data scientist)

Importantly, the model does not account for any 'insulating factors' but will only account for 'negative' data, or information that might increase risk, such as school attendance, records of domestic abuse, etc. This means that it relies on the professional case worker to account for other information or more contextual data not captured in the model, such as the person being actively engaged with social groups despite not attending schools or having a strong protective network around them outside the immediate family. As explained in one of our interviews, 'that's where we say it's down to the lead professional to say it looks like this person is at risk and they're very similar to one of our victims from last year but you know them better than we do, now make an informed decision around what you're going to do on that.' (data scientist) They went on to note, 'there's only so far you can go with data, I think' (data scientist).

Deployment and Uses

In interviews it was stated that about 450 frontline workers have access to the Integrated Analytical Hub with about 150 regular users, using different bits of the system. For the data warehouse, the system provides an overview of a family based on what is known about them across the public sector, where the data comes from and who is working with them: 'the idea is to try and break down some barriers.' (data scientist) The First Response team is using the data warehouse for deciding where to send families who engage with services, based on the overview provided in the database: 'we've got far too much traffic coming through which is overwhelming and if we can work out a way in which better decision-making could be made, then we will use our resources more wisely.' (manager) This has required 'a lot of work on trying to understand what a family actually is and how to group people together and understanding their needs.' (data scientist) It has also required that information is shared across public services. A prominent area where this is not happening in Bristol is in relation to health data, although it was noted in an interview that this is changing as health professionals are becoming more interested in social issues for health-related concerns.

In terms of the predictive modeling aspect for CSE, it was noted in interviews that this is more commonly referred to as 'targeted interventions and targeted risk assessments': 'Quite often I don't talk about predictive analytics…People get a little bit uncomfortable about the word

predictive.' (manager) Emphasised in both documents and in the interviews is the need for continued judgement from a lead professional in using predictive modeling and interpreting risk scores. One documented obtained from an FOI request stated:

> Predictive analytics should be interpreted intelligently, the results of a model do not replace a lead professional's assessment nor are the outcomes generated guaranteed. Instead the results are meant to be used as a tool to get ahead of the curve, this use of data supports an early intervention approach.

This was supported in interviews: 'it's not computer says yes or no, it's computer provides advice and guidance to the professional who adds the professional judgement in order to make better decisions about resource allocation.' (manager) Workshops and one-to-one sessions, 'upskilling staff to understand what [the model] means and how they can use it' (data scientist) was highlighted as an important aspect of the training for implementing the system amongst frontline staff. Especially as it was explained that not many people have knowledge of how the model works or the methodology of scoring. It was, however, noted that the system provides a 'context paragraph' next to the score: 'So they'll see this person is scored whatever and then they get the reason why.' (data scientist)

In terms of how the system is acted upon, a document obtained through an FOI request states:

> Using the model to identify children and young people with the most heightened risk scores, we have allocated 243 cases to key workers over the last year. As a result these families have received targeted support.

In interviews it was explained that if a child or young person has a named case coordinator, information will be sent to them that the individual has been flagged in the system, and will request information on what activity is being done to safeguard the individual, and for this to be recorded in the case notes. In cases where the child or young person is not currently receiving any support at all, someone from the commissioning team would be encouraged to 'go out and try and a proactive piece and engage that family.' (data scientist) The level of risk will also inform what team should be dealing with the case. As explained in an interview, 'If you identify, say, high, medium and low risk, we've got different services that deal with high, medium and low. So our really high safeguarding needs will go straight into our social work teams. Where there's child protection issues, we've also got a targeted youth contact which is about helping young people, signposting, self-esteem, confidence, lower level social issues and you would find your way to the right service more quickly.' (manager)

Whilst the predictive model has attempted to focus on CSE, it was noted in interviews that the Council is moving towards a wider focus that looks at exploitation more generally, creating effectively a 'vulnerability index': 'What I don't think it actually tells us is that vulnerability will play out in that you will be sexually exploited because it may be that you will be criminally exploited or you will go onto offend or you will become a drug addict.' (manager) The aim is to rebuild the model to take account of these different forms of vulnerability to create a kind

of ‘spidergram’ outlining ‘a vulnerability index with a propensity towards one thing or the other.’ (manager)

In interviews it was noted that the model is now also being used for higher-level analysis, identifying trends, patterns, geographical areas, to identify where the most risk is held. An example given was to try and identify what schools have the greatest need by aggregating all the risk scores produced: ‘we do everything at a person level and I think that’s the only way it has any value, but in saying that, once you’ve got the fine granular level, then you can always add back up again.’ (data scientist)

Auditing and Safeguards

The database is updated every week with an accompanying risk analysis. An accuracy measure will be generated each week, comparing those individuals in the target cohort to those the model has flagged. This measure will be subject to a particular threshold of accuracy: ‘what it does is it automatically tells you and alerts us once [the accuracy measure] drops below a certain threshold and says it’s no longer as good a predictor as it was six months ago, you need to rebuild it.’ (data scientist) The accuracy measure is based on a combination of ‘precision’ and ‘recall’. In documents obtained from an FOI request, these methods were outlined as:

Precision: 69% - Of those people identified by the model, 69% of these were from our target cohort.

Recall: 94% - Of our target cohort, the model identified 94% of them

Use of the system is recorded, including capturing key touches to audit what has been done on the system. This includes a trigger mechanism on volume of searches that requires a response to explain why the search has been done.

In terms of consultation or evaluation of how the system informs decision-making by frontline staff, this has not been done. This was explained in interviews as being to do with resources and ability. There also has not been any consultation with service users or an assessment of the effectiveness of targeted and early intervention. Such an assessment, it was noted, would be based on ‘re-referral rates’ (data scientist).

The data used has been shared using ‘other arms of the Data Protection Act where we’ve got statutory duty’ rather than consent: ‘we put together all of this statutory legislation that places a duty upon us and having checked all that through, we’re comfortable that in order to fulfill those statutory duties to a reasonable level, we can share this information.’ (manager)

Challenges

Several challenges with implementing and using the system was noted in interviews. They predominantly concern technical and cultural challenges. From a technical perspective, two key challenges that were emphasised are data quality and noise. Some data-sets have a high volume of errors, for example those relating to arrest records ‘with people giving wrong names, wrong date of birth, things like that’ (data scientist). There have been attempts to

account for that by weighting more reliable data-sets more highly in the system, for example information on unemployment from the Department for Works and Pensions or housing applications where people are more likely to provide correct information. Yet data can also have been recorded incorrectly and, it was noted in an interview, mistakes in the host system of any given data-set will reverberate all the way through the system. This is a problem as all the different data-sets are owned by different parts of the Council and cannot be corrected by the team behind the Integrated Analytical Hub: 'we can see data quality problems, we can't fix them. We don't own any of the systems that we report from.' (data scientist) The other concern that was expressed is a lack of filtering of data. The system is based on collecting as much data as possible and cleaned data cannot be passed back to the original data-set used as a source: 'everything we do is pulling, we don't ever push back.' (data scientist)

The other key type of challenges concerns the culture of the workplace and of the professionals working in social services. In the first instance, this concerned the challenge of getting workers to use the system, where trepidation was explained as being in part to do with a lack of technical skills and a skepticism about changing the knowledge production regarding individuals: 'There's been a strongly held view that the only people who should tell you something about them is children and families themselves.' (manager)

In terms of those using the system, the main challenge expressed in interviews concerns the nature of interpretation and change of work practices that come about from using the data warehouse and the predictive model:

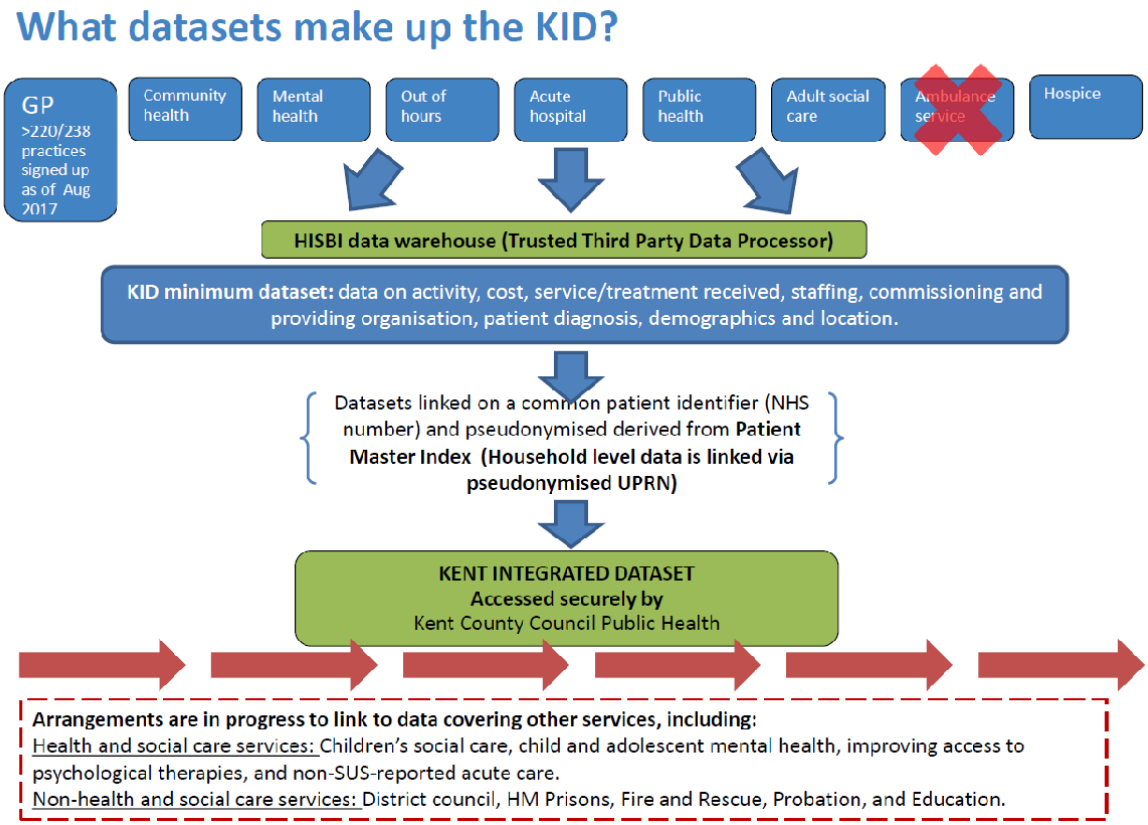
> we can't control what people do off the back of it…they might misconstrue, over-worry. It might force them into activity they wouldn't otherwise do and we don't want to generate a whole lot of concerns and worry and stress that doesn't really need to be there. (data scientist)

This was further explained as being in part to do with how risk scores are presented to workers: 'I think there's a risk that once something's in a system and somebody sees it and it's got a number next to it, they think it must be right because the computer's told me that and then they just forget all of their professional knowledge and judgement and say the computer says this.' (data scientist) There have been attempts to try and address this by deliberately not using colours like red, amber and green, or to name something 'high risk' on the system. Trying to address interpretation issues is seen as very important as a way to mitigate the risk of feedback loops emerging from actions taken on scores (for example, alerting police because of a high risk score, in which contact with the police is a factor for generating a higher risk score): 'You've got to be careful you don't end up generating some feedback loops where your scores feeds another score feeds your score, and you end up just constantly multiplying everybody's score each week. There's definitely a risk of that.' (data scientist)

Related to this is the challenge of shifting to 'preventative proactive work' amongst workers who are 'used to responding to high levels of need in crisis' (data scientist). 'Capturing more risk' through the use of automated risk assessments will require engagement with individuals who are not usually considered high enough need, asking for a 'light touch' that engages with

people on an on-going basis. Expanding the pool of risk through automated risk assessments also introduces issues in a context of austerity and limited resources. At one level, this was noted as a reason for some hesitancy around the implementation amongst some workers: 'There's been elements of why do you want to go and work out the risk and vulnerability because we've already got loads of people that are coming through the door with risk and vulnerabilities and what's the point of finding even more people?' (manager) On another level, it also introduces ethical challenges with regards to how you then decide whom to work with: 'we have an internal ethics challenge board where we submitted some questions to them to say I've got a list of 100 people, I can only work with ten, how do I pick the right ones? Or what do we do with the ones that don't receive any support because we've identified that there is a need there but we can't resource it, so what should we do?' (data scientist) In interviews it was noted that the approach so far has been based on being able to provide a justification for the decision to work with one person and not another.

Kent Integrated Dataset



Source:
https://www.local.gov.uk/sites/default/files/documents/W5.%20Shifting%20the%20focus%20to%20prevention%20and%20early%20intervention%20-%20Dr%20Abraham%20George.pdf

Summary

The Kent Integrated Dataset is used for population health planning. It is not used to make decisions about specific individuals. Data about individuals in the dataset are pseudonymised.[19] Access to the dataset is limited to the Kent Public Health Observatory team. Controlled access is possible for research purposes but requires application. In the spring of 2018 Kent County Council noted that they have not deployed predictive analytics processes but that they have hopes to in the future. They have done analyses that 'project into the future based on past trends'. They are developing a system dynamics model, detailed below, to see the impact of health prevention efforts on reducing smoking, obesity and others.The Kent Clinical Commissioning Groups have contracted Optum, a healthcare intelligence company, to provide their business analytics service. Optum is going to create a new Kent Integrated Dataset to inform commissioning decisions. Optum is part of the American UnitedHealth Group. More research into the development of this new Optum integrated dataset is ongoing. We requested an interview with Optum to find out more details about the new dataset, but an Optum representative declined.

[19]You can find more information about pseudonymisation here: https://www.i-scoop.eu/gdpr/pseudonymization/

The details provided in this summary concern only the first Kent Integrated Dataset developed and controlled by Kent County Council and not details about the new Optum system.

Implementation

The Kent Integrated Dataset has been described as following the model of Welsh and Scottish whole population electronic healthcare records, but goes further by including data 'from a wider range of health and care services'. It also provides data from the entirety of a local population.[20] The dataset comprises the data of almost two million residents living in South East England.[21] The goal is to enable Kent and Medway planners to know more about how people are using their services, the services needed and also how this connects to wider socioeconomic and environmental contexts influencing health and service use.[22] The dataset includes data from 2014 to the present and is continually updated.

Enabling legislation – for Public Health

Date	Name of legislation	Requirement of Public Health departments
2002	The Health Service (Control of Patient Information) Regulations Act	Grants Public Health teams access to confidential patient information to, amongst other things, recognise trends in diseases and risks.
2006	NHS Act (as amended)	Included adult social care users to confidential patient information access rights in 2002 Act.
2007	Local Government and Public Involvement in Health Act	Local authorities are required to produce a Joint Strategic Needs Assessment of the health and well being of their local community
2012	Health and Social Act	Major reorganisation of NHS services.
2016	General Data Protection Regulation	The new data protection act.

Source: The Kent Integrated Dataset (2017) Presentation, August, available: https://www.kpho.org.uk/__data/assets/pdf_file/0004/74146/Kent-Integrated-Dataset-August-2017.pdf

As indicated in the table above, there were pre-existing pieces of legislation that enabled Public Health to access data but that also protected citizen data. The 2007 Local Government and Public Involvement Act and specifically the requirement that local authorities produce Joint Strategic Needs Assessments of the health and well-being of their communities was a key stimulator for the Kent Integrated Dataset. Previously those in Public Health could access

[20] Lewer, D., Bourne, T., George, A., Abi-Aad, G., Taylor, C. and George, J. (2018) Data Resource: the Kent Integrated Dataset (KID), *International Journal of Population Data Science*, 3(6): 1-8.

[21] Lewer et al. 2018

[22] Lewer et al. 2018, p.2.

data about single conditions, but there was recognition that being able to link up more datasets would provide a better picture of community health needs and service impact.

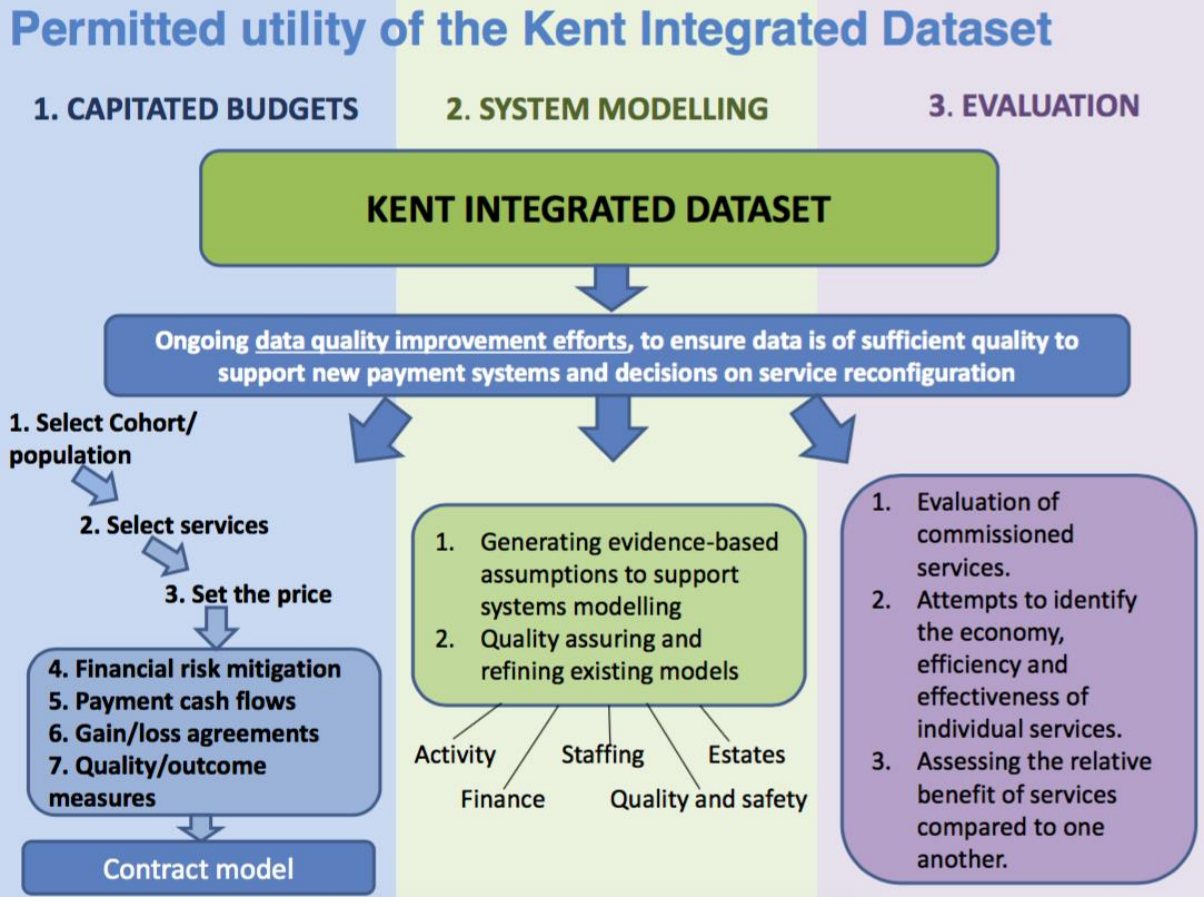
Several contextual factors are important: a) increasing research evidence about the complex relationships between social and environmental contexts and health needs, b) local authorities facing cuts to funding that has influenced their ability to provide social care services and increased the needs of their citizens, c) and a national NHS funding gap influencing health care.

The rationale for developing an integrated dataset was detailed in 2013. It has been argued that those commissioning services need to focus more on prevention and to do so by commissioning more integrated services and making better use of intelligence systems':

> 'Commissioners need the relevant resources and technical expertise to develop a longitudinal system using metrics that are person centered / population based, rather than the activity or performance of individual organisations or services. Improving the health and wellbeing of the population requires commissioners to have a cross sectional understanding of how prevention and preventative services impact differently at different population risk groups….
>
> 'This paper makes a case for whole systems intelligence and a need to have a cultural shift from analysing data at an organisational level to analysing information across the complete patient pathway. This should include health and social care as well as information on socio, economic and environmental factors that contribute to health and wellbeing. In this regard it is about the effective sharing and management of information at a citizen level, scaled up to a population level to effectively understand the holistic nature of integrated care and the many confounding factors that affect health and social care outcomes and a person's resilience to improved wellbeing.'[23]

By linking up the datasets, the benefit has been the ability to better understand the public health needs of the people in Kent and Medway, service use and the impact of services. Before the Kent Integrated Dataset, healthcare and social care could only be looked at in isolation and people could only look at data relating to single conditions.

[23] Gough, R. (2013) 'Integrated Intelligence: how will it support integrated commissioning', available: https://democracy.kent.gov.uk/documents/s43602/Item%207%20b%20HWBB%20integrated%20intelligence.pdf



Source:
https://www.kpho.org.uk/__data/assets/pdf_file/0004/74146/Kent-Integrated-Dataset-August-2017.pdf

In a presentation the Kent Integrated Dataset is described as having its use restricted to four purposes:[24]

1) To assess return on investment by providing feedback on commissioning decisions, to determine the impact of different services, and to determine the impact of non-NHS, public sector, services on health and well-being (such as housing, education, police)
2) To provide evidence that can be used for service design
3) To assist with the design of payment models and support new models of care
4) For public health intelligence by supporting analytics that can enable effective prevention targeting and population budgeting

The strengths of the Kent Integrated Dataset are said to be[25]:

- That it covers areas other electronic health records don't such as community health, mental health and social care
- That it includes new variables, specifically the cost of an episode to enable economic modelling

[24] The Kent Integrated Dataset (2017) Presentation, August, available: https://www.kpho.org.uk/__data/assets/pdf_file/0004/74146/Kent-Integrated-Dataset-August-2017.pdf

[25] Lewer et al. 2018, p.6.

- That the dataset is updated regularly so it enables quick evaluation of service changes
- That the use of a unique reference number for each user means patients can be tracked across services so researchers can get a better picture of the paths they take

There are a wide range of presentation videos and power point slides as well as publications available online detailing what the integrated dataset involves, the rationale for its use and some examples of how it has been used. Some examples of particular uses have been detailed below.

There is little information available about the new Optum integrated dataset. Also unclear is how the governance of the new dataset may differ from the one in existence now. Research into this is ongoing.

Model

The Kent Integrated Dataset is an attempt to 'provide a single dataset across all Kent public services and put together by the Kent & Medway Clinical Commissioning Groups and Kent County Council Public Health. The dataset brings 'together data from 250 local health and social care provider organisations" as well as Fire and Rescue Service data to support planning and commissioning decisions. It is referred to as an early linked dataset initiative and 'possibly the largest linked dataset of its kind'.[26]

What information does the KID hold?

Demographics	Segmentation tools	Admin	Diagnoses	Activity/cost	Service
Age	IMD	Practice code	Morbidity profile (Read codes)	Contact date	Healthcare Resource Groups (acute)
Sex	CPM (Risk Stratification tools)	Provider code	Referral source	Cost/price	Tariff cluster (mental health)
Lower Super Output Area	MOSAIC	Commissioner code		Point of delivery	Care Package (social care)
	ACORN				Service code (community)
	eFI (Frailty score)				Specialty (outpatient)
	ACG (Restricted use)				Staff type

[26] Oakford, P., Scott-Clark, A., Godfrey, V., Whittle, D. (2017) The Kent Integrated Dataset (KID), Health Reform and Public Health Cabinet Committee, 30 June, available: https://democracy.kent.gov.uk/documents/s77422/item%2013%20-%20KID%20report.pdf

Datasets not included (as of Jan. 2018) are: sexual health data, suicides, children's social care data, hospital care not funded through NHS and the interactions of residents with care providers outside of Kent and Medway.[27] It was agreed in principle that data from Specialist Children's Services and education data would flow into the KID.[28] Kent Police have agreed in principle that they are ready to share their data.

The system works through pseudonymisation.[29] An encrypted version of people's unique NHS identifier number is used to link the records of individuals across the different datasets including general practices, hospitals, community health services and social care. Pseudonymization happens at the source of where the data is generated. 'Names are excluded and other potentially identifiable information is coarsened to prevent re-identification of individuals. For example, dates of birth are replaced by single-year-of-age and postcodes are replaced by Lower Super Output Areas (a geo-graphical area covering approximately 1500 residents).'[30]

The system was developed in-house with start-up funds from the NHS 'National Long Term Conditions Year of Care' programme. When that funding ended Kent County Council and the local National Health Service agreed to continue funding and development.[31]

'The KID comprises individual-level linked EHRs from the following services located in Kent and Medway: primary care providers (including general practices, out-of-hours providers and walk-in centres), community health providers, mental health services, acute hospitals (including accident and emergency, inpatient and outpatient episodes), public health services, adult social care, fire & rescue service and palliative care hospices. The dataset includes records of interactions between residents of Kent and Medway and these services. Each service provider/data owner has securely uploaded data monthly since April 2014.'[32] Both Acorn and Experian socio-economic profiling tools are used as part of the KID.[33] The Maidstone and Tunbridge Wells NHS Trust hold the contracts for extracting GP data and linking the KID to the CACI ACORN and Experian Mosaic population segmentation systems.[34]

The kinds of data included vary by type of service used and event. Kent says that a full set of the variables used is available on request.[35]

---

[27] Lewer et al., 2018, p. 6.

[28] Oakford et al. 2017.

[29] For more information on pseudonymisation see: https://www.i-scoop.eu/gdpr/pseudonymization/

[30] Lewer et al. 2018, p.2.

[31] Lewer et al. 2018, p.2.

[32] Lewer et al., 2018, p. 2.

[33] Details of both are available here: https://acorn.caci.co.uk/ and here: https://www.experian.co.uk/marketing-services/products/mosaic-uk.html

[34] Kent County Council (2018) Response to Data Justice Lab Freedom of Information Request, FOI Reference 1185536, 9 March.

[35] Lewer et al., 2018, p.2.

Deployment and uses

The main user of the Kent Integrated Dataset is Public Health. Analysis is done for the Director of Public Health and to also help the CCG's develop the Joint Strategic Needs Assessment. An example provided was that the KID helps those in public health assess and respond to complex issues like multimorbidity. One of the main impacts noted was a shift toward greater focus on prevention and the value of preventive action.

> [W]e are a public health team doing population health analysis, we don't need to access patient identifiable data because that's not our interest. Our interest is population level or population health analytics and most of our work is done on that basis (manager).

The dataset has been used in different ways to better target information to people, better understand services and how they may need to be improved. Work identifying the frail and elderly has led to the development of local care models that emphasize care for this group and the development of services to target this group.

Another example is work done with Public Health England in relation to health check equity audits. The programme looks at people between 40-74 who have a cardiovascular risk that has not been identified yet in order to prevent harm down the road. Through analysis they identified that certain segments of the population were less likely to go for a check-up and this information is being used to change the way the health checks programme is delivered to reach those groups.

The Public Health Observatory have undertaken an evaluation of Kent Fire and Rescue Service's Safe and Well visits using the KID data. The evaluation seeks to identify if the visits have a positive impact on health outcomes, for example by reducing the incidents of hospital visits compared with a matched cohort.[36] The initial evaluation has not identified a significant impact on health outcomes although it established the methodology in order that the affect can be analysed over the next three years.

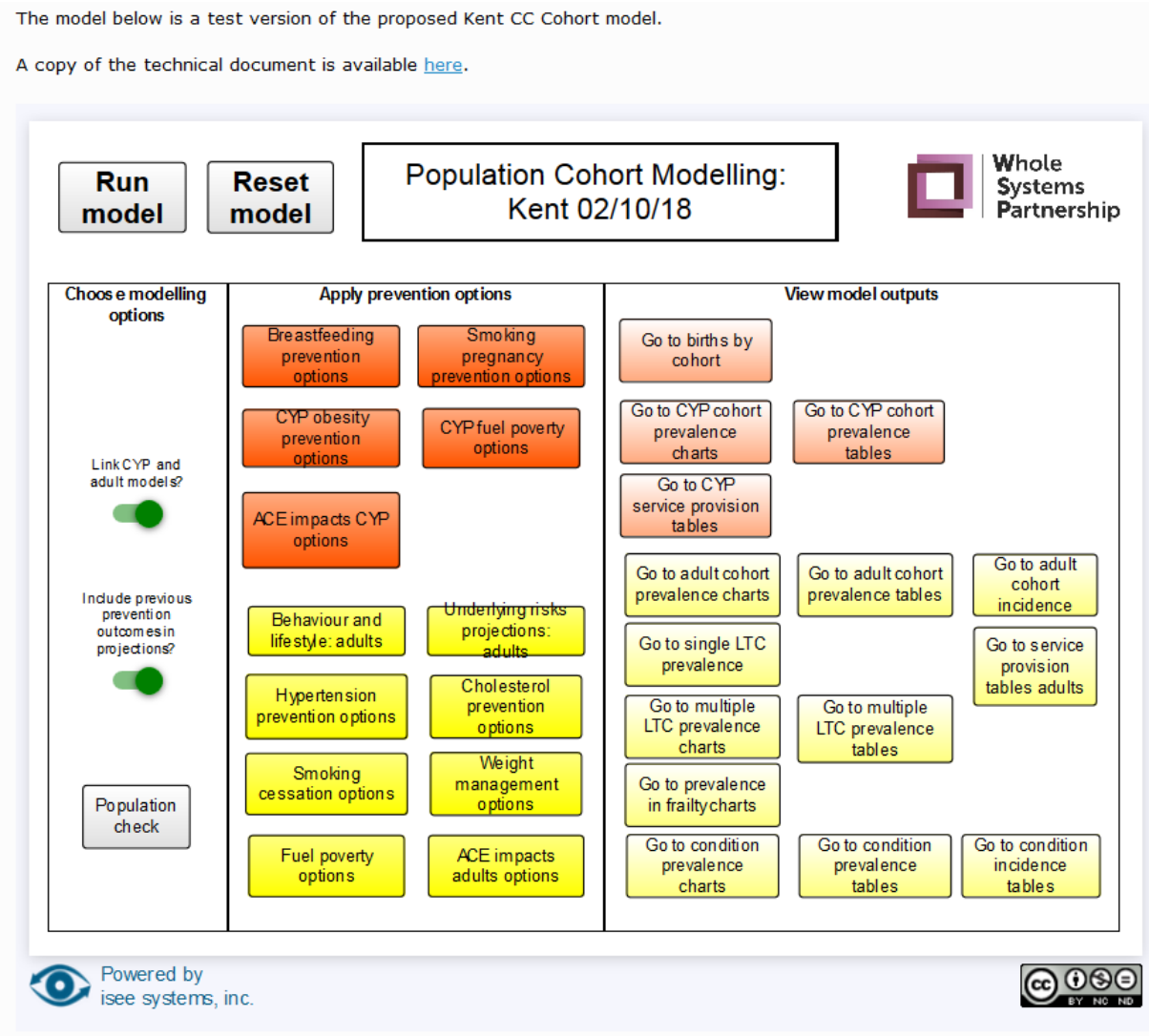
Kent Fire and Rescue Service is also talking with the Public Health Observatory about how they can use analysis to identify risk factors, like poor health, that make some groups more at risk of fire. The goal is to use this to target preventation such as Safe and Well visits.

The JSNA population cohort model 'seeks to transform the Kent JSNA into a forward planning commissioning tool'.[37]

---

[36] Kent and Medway Fire and Rescue Authority (2018) Customer and Corporate Plan 2018-2022, available: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=2ahUKEwjWubnDufTeAhUBZlAKHUSQB-kQFjAAegQIAxAC&url=http%3A%2F%2Fwww.kent.fire-uk.org%2FEasySiteWeb%2FGatewayLink.aspx%3FalId%3D16221&usg=AOvVaw2NQaACejigWZ_ncRV2W6Ta

[37] Kent Public Health Observatory (2018) JSNA Population Cohort Model, available: https://www.kpho.org.uk/joint-strategic-needs-assessment/jsna-population-cohort-model



The tool uses "systems dynamics" to 'model possible impacts of key policy and service capacity changes'.[38] It's said that the model can be used to 'test "what if" scenarios focusing on additional investment on prevention':

> The model seeks to integrate and synthesise best estimates from a variety of sources to estimate the extent to which a range of factors, acting in combination, explain or predict certain health outcomes. Key model outputs include projected incidence, prevalence of long term conditions as well as population cohorts relevant for the Kent & Medway STP, including, for example, those affected by adverse childhood experience (ACE) as it is modelled into later life.[39]

[38] Kent Public Health Observatory (2018) JSNA Population cohort model, available: https://www.kpho.org.uk/joint-strategic-needs-assessment/jsna-population-cohort-model

[39] Ibid.

Auditing and safeguards

Only the public health observatory team within Kent County Council has regular access the integrated dataset and access is restricted to a controlled location. This has been connected to addressing people's concern about meeting the Data Protection Act and obligations around confidentiality.

Kent County Council has allowed 6 analysts from NHS England to access the KID via contractual agreement. Access has also been provided to Carnall Farrar (private consultancy) and Whole Systems Partnership.[40]

It is possible for researchers to get access to the KID by application for research as long as 'the research is likely to provide some benefit to the Kent and Medway health and care economy'.[41] It is said that to date the data has been used mostly for healthcare planning and not for research purposes. It has been noted that a number of universities want to work with the KID.

An NHS Data Sharing Audit was done in May 2017 at Kent County Council Public Health Intelligence.[42] The audit measured how uses of data conforms to the data sharing framework contract and the data sharing agreement in relation to the Office of National Statistics birth and death data. The audit also involved investigating if practices conform to Kent's own policies and procedures. The audit team reported that 'there is low risk of a breach of information security, duties of care, confidentiality or integrity'.

Two privacy impact assessments have been done. As part of the General Data Protection Regulation, a Data Protection Impact Assessment will need to be done on anything that is new or added to the KID. The Council has a GDPR working group to ensure compliance.

'Data owners are responsible for validating and checking the quality of data before it is fed into the KID'... After each monthly upload to the KID, the data owners check that the correct total number of records is registered in the KID. The KID team then runs five checks on each 'service function' (primary care, social care, hospitals, etc.) to monitor data quality'.[43]

'Currently, individual-level data can only be viewed and analysed on Kent County Council's computer systems, with access provided physically at Kent County Council or via a secure remote desktop'.[44]

---

[40] Kent County Council (2018) Response to Data Justice Lab Freedom of Information Request, FOI Reference 1185536, 9 March.

[41] Lewer et al. 2018, p.7

[42] NHS Digital (2017) Data Sharing Audits, 20 July, available: http://webarchive.nationalarchives.gov.uk/20180328130852tf_/http://content.digital.nhs.uk/media/24796/Data-Sharing-Agreement-Audit--Kent-County-Council/pdf/Data_Sharing_Agreement_Audit_Report_-_Kent_County_Council.pdf/

[43] Lewer et al., 2018, p.2.

[44] Lewer et al., 2018, p.7.

It is noted that the GDPR has specific requirements in terms of documentation including and related to data protection, impact assessment and IT data protection toolkit.

Confidentiality is maintained through pseudonymisation. Data is required to be anonymised before being published and its use is said to conform to the Information Commissioner Office's anonymisation code. Data controlling organizations, such as a GP practice or community organization that provides data, are brought together as part of a board that governs the Kent Integrated Dataset. If people want to do a piece of analysis they bring it before the board and anyone on the board can stop an analysis they are not comfortable with.[45]

The data warehouse audits the queries that are run and any overly specific search it is said should ring an alarm.

In terms of data quality, data quality issues are noted in report footnotes. Someone at the Observatory is responsible for producing a monthly data quality summary that indicates issues. The Observatory keeps a data quality matrix that flags every issue identified, split by provider type. Also users of the data feed back any data quality issues they find to ensure they are addressed.

KCC is now working through how governance arrangements of the KID may need to be changed to promote research and development needs. As part of this the General Data Protection Regulation is influencing decisions made:

> 'The GDPR emphasises the importance of having a contractual arrangement to actually underpin any data sharing arrangement. So we have to be clear about the customer / supplier arrangement and control set. How does a customer instruct the supplier to actually process the data in the right way? Within those governance arrangements we would have ideally a research ethics committee which will be the robust process for managing any sort of research requests… So the plan in principle is to try and use existing forums, advisory public and patient engagement groups to be as part of the governance for the KID going forward' (manager)

It has been stated that many local authorities are looking to copy the Kent model.[46]

System details

'The KID was built from existing systems, using a SQL-server data warehouse (The Kent and Medway Health Informatics Services) and a purchased business intelligence tool. KMHIS developed a 'black box algorithm' for data as it enters the KID. This runs an algorithm to check that the information on either side matches up, without the system operators needing to see identifiable data. The programme then creates a de-personalised NHS number – a 64 bit

[45] The Kent Integrated Dataset (2017) Presentation, August, available: https://www.kpho.org.uk/__data/assets/pdf_file/0004/74146/Kent-Integrated-Dataset-August-2017.pdf

[46] Kent County Council Health Reform and Public Health Cabinet Committee (2017) Agenda and minutes, 30 June, available: https://democracy.kent.gov.uk/ieListDocuments.aspx?CId=895&MId=7736&Ver=4

sequence – which is given to each piece of personal data and is applied to all files that come in relating to the same person. The project was assigned an IG expert who advised on the requirements of KID and informed the data sharing agreement for each agency, including the 240 GP surgeries in Kent'[47]

Kent County Council uses Microsoft SQL server Management Studio version 17 to access data held in the Kent Integrated Dataset. Data is extracted into three statistical software programmes for analysis: SPSS, STATA, Excel.[48]

Patients can opt-out by letting their GP know they do not want their data to be shared and used, although in a 2017 committee meeting it was suggested that there may be changes to this process.[49] As of January 2018 it was reported that 2.3% of patients had opted out from sharing their data. As of December 2017 it was reported that 93 percent of primary care providers were sharing their data.[50]

Risk assessment tools deployed in the Kent Integrated Dataset include the King's Fund model used to identify the risk of unplanned admittance to hospital. It is noted that this is 'a well-established risk assessment tool in the NHS'[51] It has also been noted that Kent County Council uses an electronic frailty index that has been detailed by the NHS.[52] The Index uses health record data to identify and grade the severity of frailty. The goal is 'to enable treatments and services to be targeted to a person based on their frailty status rather than their chronological age'.[53]

Challenges

Minutes of meetings from the Kent County Council Health Reform and Public Health Cabinet Committee show that members have raised concerns about the potential for the data to be vulnerable to cyberattack. The response has been that the Council continually mitigates against this and complies with Information Governance Standards and NHS Digital guidelines for data security.[54]

---

[47] Involve, 'Example A: Kent Integrated Dataset (KID)', available: https://www.involve.org.uk/sites/default/files/field/attachemnt/workshop-examples.pdf

[48] Kent County Council (2018) Response to Data Justice Lab Freedom of Information Request, FOI Reference 1185536, 9 March.

[49] Kent County Council Health Reform and Public Health Cabinet Committee (2017) Agenda and minutes, 30 June, available: https://democracy.kent.gov.uk/ieListDocuments.aspx?CId=895&MId=7736&Ver=4

[50] Lewer et al., 2018, p.6.

[51] Kings Fund (2007) Predicting and reducing re-admission to hospital, available: http://clahrc-yh.nihr.ac.uk/our-themes/primary-care-based-management-of-frailty-in-older-people/projects/development-of-an-electronic-frailty-index-efihttps://www.kingsfund.org.uk/projects/predicting-and-reducing-re-admission-hospital

[52] NHS Development of an electronic Frailty Index, available: http://clahrc-yh.nihr.ac.uk/our-themes/primary-care-based-management-of-frailty-in-older-people/projects/development-of-an-electronic-frailty-index-efi

[53] NHS Frailty Index.

[54] Kent County Council Health Reform and Public Health Cabinet Committee (2017) Agenda and minutes, 30 June, available: https://democracy.kent.gov.uk/ieListDocuments.aspx?CId=895&MId=7736&Ver=4

Other committee members raise concerns about the potential to re-identify people, to this the response is that the County Council does not have 'the scope to re-identify pseudonymised data collated by the NHS and would only ever have access to the pseudonymised version'[55]

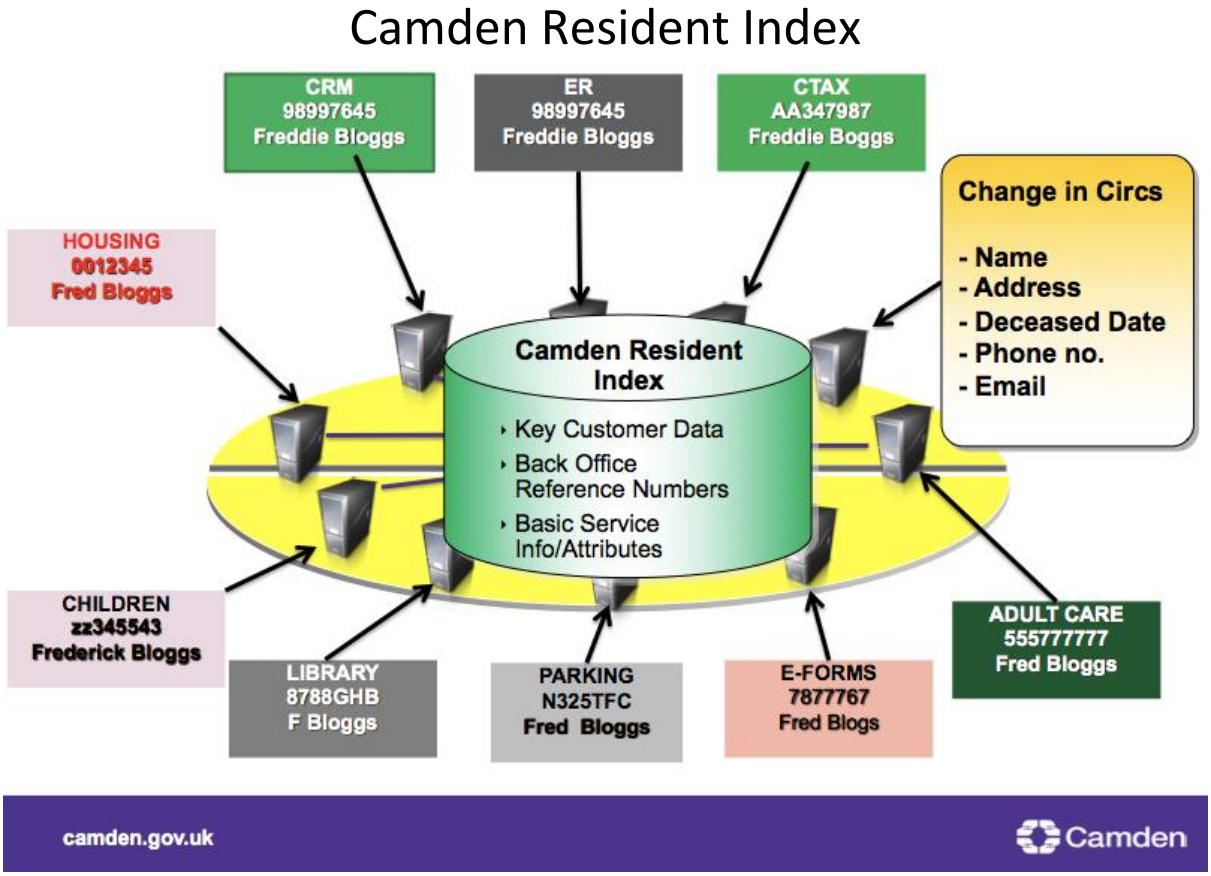
Data quality is an ongoing issue with quality varying across datasets. Also noted is that data are not always recorded consistently.

There is ongoing concern about how to engage and inform the public about data uses. Those who developed and use the KID are in the process of determining how to better consult with the public about uses of data in a way that engages and leads to greater understanding.

The GDPR is requiring those using personal data, even if it is pseudonymised, to be 'more precise and strict in the design, implementation and enforcement' of their code of practice'.[56]

[55] Kent County Council Health Reform and Public Health Cabinet Committee (2017) Agenda and minutes, 30 June, available: https://democracy.kent.gov.uk/ieListDocuments.aspx?CId=895&MId=7736&Ver=4

[56] George, Abraham (2017) Kent Integrated Dataset: Use of linked data for applied analytics to support service planning', Presentation, August, available: https://www.local.gov.uk/sites/default/files/documents/W5.%20Shifting%20the%20focus%20to%20prevention%20and%20early%20intervention%20-%20Dr%20Abraham%20George.pdf



Summary

The Camden Resident Index is a data management system utilising software supplied by IBM that allows for a 'single view of a citizen' by aggregating data from 16 different council business systems across Camden Council, covering 123 fields of primarily demographic information. It is to date the largest data management installation in local government in the UK.[57] It was created in 2013 following the closure of the national children's database Contact Point to uphold multi-agency work. It uses probabilistic matching technology to match individuals or households across the different business systems, in which records are matched together to produce a comparison score that indicates the likelihood of records belonging to the same person or family. The Camden Resident Index is used by the Multi-Agency Safeguarding Hub to locate information about a household's engagement with services across the Council. A key use of the index is to enable fraud detection, such as validation for residency for accessing council services such as school places, number of residents in a household for council tax discount, or cases of illegal subletting of council housing.

Implementation

Camden Council is one of the first local authorities to have implemented a master data management system, 'to allow the single view of a citizen.' (business analyst) The rationale

[57] https://www.involve.org.uk/sites/default/files/field/attachemnt/3-case-studies-data-sharing.pdf

for implementation came from both a view 'to improve customer service' and 'the frustration of residents' (business analyst) with different parts of the council not being aware of previous engagement. It was also reasoned as a business case for improving fraud detection. It followed on from previous attempts at integrated data sets in Children's Services which were closed following a change in government: 'Contact Point was a national children's data base which was set up but soon after it was launched, there was a change in Government and it was scrapped but I think people had got a taste for how multiagency working, different agencies could get a single view of a child.' (project manager)

The decision to contract IBM as supplier followed a proof of concept stage: 'we ran...eight datasets in a snapshot dataset, we looked at the end result, we then identified potential savings, particularly through fraud, from that and the business case was made and we carried out a standard procurement. Two or three companies bid and IBM won.' (former employee) Whilst the software, which is similar to anti-fraud technology used in the banking sector (former employee), is supplied by IBM, the data model is accessible to the Council and the data that informs the matching process and how it is weighted can be adjusted: 'we have control over the matching algorithm and over the last year we have regularly reviewed the algorithm. We are getting good results in terms of matching residents.' (business analyst)

Model

The Camden Resident Index uses probabilistic matching technology 'which is machine learning decision making on 16 of the Council's biggest databases.' (former employee). According to an FOI request processed in November 2017, data sources used for the Camden Resident Index include:

- Customer transactions
- Housing
- Council Tax and Benefits
- Electoral Register
- Adult and Children's Social Services
- Schools and pupil information
- Parking Control & permits, accessible transport
- Young people's information
- Libraries[58]

Using the IBM software, records from different departments are matched to determine how likely it is that two records are the same person: 'two records are compared and they get a comparison score and the closer they are, the higher the score. And if things diverge, like

58 https://www.whatdotheyknow.com/request/445907/response/1083921/attach/3/FOI%20Response%20FOI10251.pdf?cookie_passthrough=1

they've got different dates of birth, the score will go down.' (project manager) This then allows for some data discrepancies between different systems. In order to overcome discrepancies, it is possible to create a 'Camden most trusted view' (project manager) which is based on attributing different levels of verification to different datasets: 'Data is verified differently in different business areas. For example, when signing up for a council tenancy, residents need to provide a passport, whereas a library service wouldn't require such stringent verification. How we form our trusted view of data is based on this detailed knowledge of business processes and we are able to configure on an ongoing basis.' (project manager)

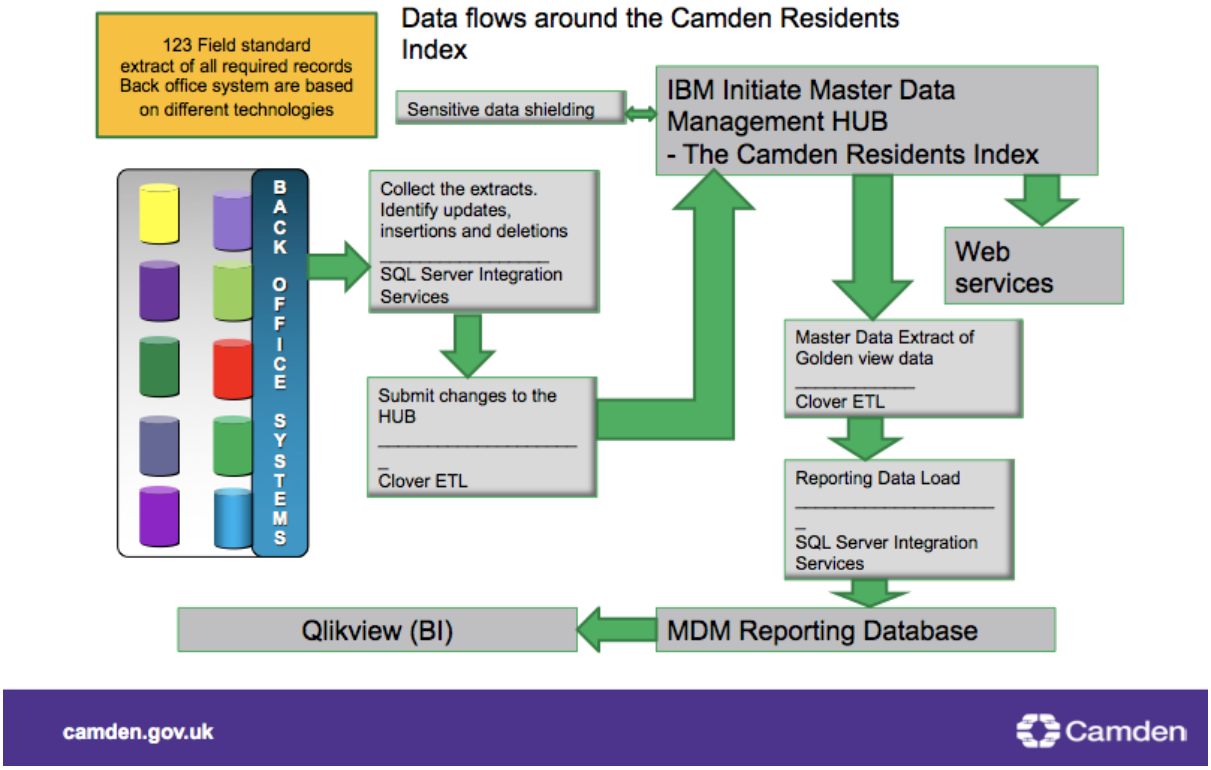
The index can also provide a household view which is pivoted around the address which joins data together to show the different record from the different people associated with that address.

FIELDS	HB	CRM	HSD	ER	ADULTS	GOLDEN
Full Name	**Fred Bloggs**	Freddie Bloggs	Frederick Blogges	Fred Bloggs	Fried Bloggs	**Fred Bloggs**
DOB	**10-1-1965**	01-01-1900	10-1-1965		10-1-1965	**10-1-1965**
Gender	M	M	M		M	**M**
Address	10 The Grange Camden NW1 0AA	Flat A 10 The Grange Camden NW1 0AA	10 The Grange Camden NW1 0AA	**Flat A 10 The Grange Camden NW1 0AA**	10 Grange Camden NW1 0AA	**Flat A 10 The Grange Camden NW1 0AA**
Home Tel		0231-987-7866	**0205-937-996**			**0205-937-996**
Mobile	**07988674707**	077134563				**07988674707**
Email		**fblogg@v.com**	bloggy@bt,com			**fblogg@v.com**
Income Support	Y					**Y**
Customer Care Ind					**Do not visit address alone**	**Do not visit address alone**

Further details about the Camden Resident Index and the workings of the model for flagging individuals and/or households as potential fraudsters were withheld in response to an FOI request on grounds of exemption.[59]

[59] https://drive.google.com/file/d/1XmCxdJnWQIivSWnWEwA-Hw3gnj8wcGbO/view

Deployment and Uses



The Camden Resident Index is used by a number of teams within Social Care, Housing, Accessible Transport, Electoral Services, Customer Services, Internal Audit, Planning.[60] According to Camden Council's privacy statement, approximately 350 staff across the council have access to the Camden Resident Index. It further notes, that 'Their access is tailored on a need to know basis and the majority of the system users will only access the Adults' records. Access to children's records is restricted to children's services staff (the complex families teams, children missing education, safeguarding, admissions fraud etc.), system administrators and audit teams.'[61]

One of the key use cases for the Camden Resident Index is fraud detection. This is particularly enabled from the household composition, for example to detect illegal subletting or benefits claims, or school admissions. Fraud detection can include 'school admissions where people are applying for school places from places they don't live in, or people illegally subletting their council tax properties, or people retaining accessible transport benefits when they no longer live in the borough or parking permits; the council services to which they're not legitimately entitled. This would give us indicators as to people who may fall into that category because

60 https://www.whatdotheyknow.com/request/445907/response/1083921/attach/3/FOI%20Response%20FOI10251.pdf?cookie_passthrough=1

61 https://www.camden.gov.uk/ccm/content/council-and-democracy/about-the-council/about-this-site/disclaimer-and-privacy-statement/?page=2

the patterns don't match the resident, they don't appear to be resident or they appear to have moved.' (project manager)

Importantly, it was noted in interviews that such fraud detection would not automatically lead to a final decision. Instead 'it helps the service whittle down the likely cases to investigate…you'd always want to rely on the human judgement based on the information.' (project manager)

Another related use of the Camden Resident Index mentioned in interviews is eligibility for services. Not in terms of evaluating someone's eligibility, but in order to eliminate the need to re-apply for services: 'so if you're thinking about Freedom Passes or disability badges, parking blue badges, that sort of thing, then in Camden what we did, using the data we had available from the 16 systems, is switch the whole thing around and say you no longer need to apply for these things because we know you're registered with adult social care, you're receiving a service as a disabled person and nothing has changed.' (former employee)

Beyond fraud and eligibility, the Camden Resident Index is used by teams such as the Multi-Agency Safeguarding Hub that get referrals about missing children or children at risk or vulnerable adults as a way to get information from different systems immediately: 'In the past they had to go into each individual system separately to find out information but now on the Camden Resident Index, they can go in and find out the information about that particular child and they can find out about the services that that child is being engaged with or is known to, and then they can also find out information about the household and their engagement with our services.' (project manager) According to an overview of the system provided by the charity Involve, this deployment of the Camden Resident Index has 'enabled frontline professionals, such as multidisciplinary social workers, to do their jobs in ways that wouldn't previously have been possible, such as flagging safeguarding issues which otherwise wouldn't have been seen by legacy systems.'[62]

Predominantly the Camden Resident Index is used for providing a view of individuals or households, but it was also mentioned in interviews that some neighbourhood and population level insights are drawn from the index, also for the purposes of predictive analytics. For example, it was noted in an interview that birth rates documented in different areas of the borough informed decisions about targeting children's centre services according to local need – which centres to retain and which to potentially decommission (former employee). More generally, future plans for the creation of a 'data lake' was mentioned in interviews, in which structured data would be combined with unstructured data, to include transactional data that can help spot patterns and inform decisions about what services are needed.

Interviewees at Camden expressed hesitancy about profiling at individual level and there is no 'scoring' practice taking place that is aimed at predicting future behaviour. However, in one interview it was noted that the Children's Services and the Troubled Families Programme have sought to create data visualisation tools that draw on the Camden Resident Index, such

[62] https://www.involve.org.uk/sites/default/files/field/attachemnt/3-case-studies-data-sharing.pdf

as the Family Support Dashboard, which includes 'family customer journey mapping to show which services over time a particular family has touched and what is the likely pathway for those families based on similar families in the past.' (former employee) In subsequent correspondence, this was clarified as being a matter of using 'linkages created in the Camden Resident Index matching engine to meet the requirements of the national Troubled Families programme. This involves the linkage of data from different council databases to identify complex need and monitor the progress of having received a service' (business analyst) and was noted as being markedly different to profiling.

Auditing and Safeguards

Primary evaluation of the creation of the Camden Resident Index concerned the business case. In an overview of the system carried out by the charity Involve, it was noted that Camden estimates that it has saved £800k just from identification of illegal subletting of council housing. In an FOI request it was noted that the budget for maintaining the Camden Resident Index is £50,000 per annum in addition to ongoing staff costs for support and development which is approximately another £50,000.[63]

Some consultation was also carried out when the system was first introduced with a 'stakeholder panel to give input' run by customer services with a team that included members of the public, 'to help shape it.' (former employee) Further consultation has not been carried out beyond the information that is available about the system in the council's privacy statement. It was noted in interviews in relation to the impact of the EU General Data Protection and Regulation Act (GDPR) that 'most of the information data that we collect in the council, we collect because of statutory duty to care or to provide services and not based on consent.' (business analyst)

Performance tuning of the matching algorithm comes from working 'closely with the Resilient Families programme [previously the Troubled Families programme].' It was explained, 'when we make changes to the algorithm they will check for duplicates and for false positives.' (project manager) but it was not explained as a systematic auditing review. There is a regular audit of uses of the system that produces logs of activity including 'how often people are searching, so if there's unusual activity, if people seem to be trawling the system, we could flag it.' (project manager)

Challenges

A key challenge with the Camden Resident Index noted in interviews is the data quality of the different business systems, either because of lack of information, or wrong date of birth, or because the format of entering information is different between different systems: 'This does shine the light on our system's data quality….Data quality is always an ongoing challenge and the best thing to do is address it at source where there's a backlog of data.' (project manager) This also means training people to input data differently or to enforce better quality data:

63 https://www.whatdotheyknow.com/request/445907/response/1083921/attach/3/FOI%20Response%20FOI10251.pdf?cookie_passthrough=1

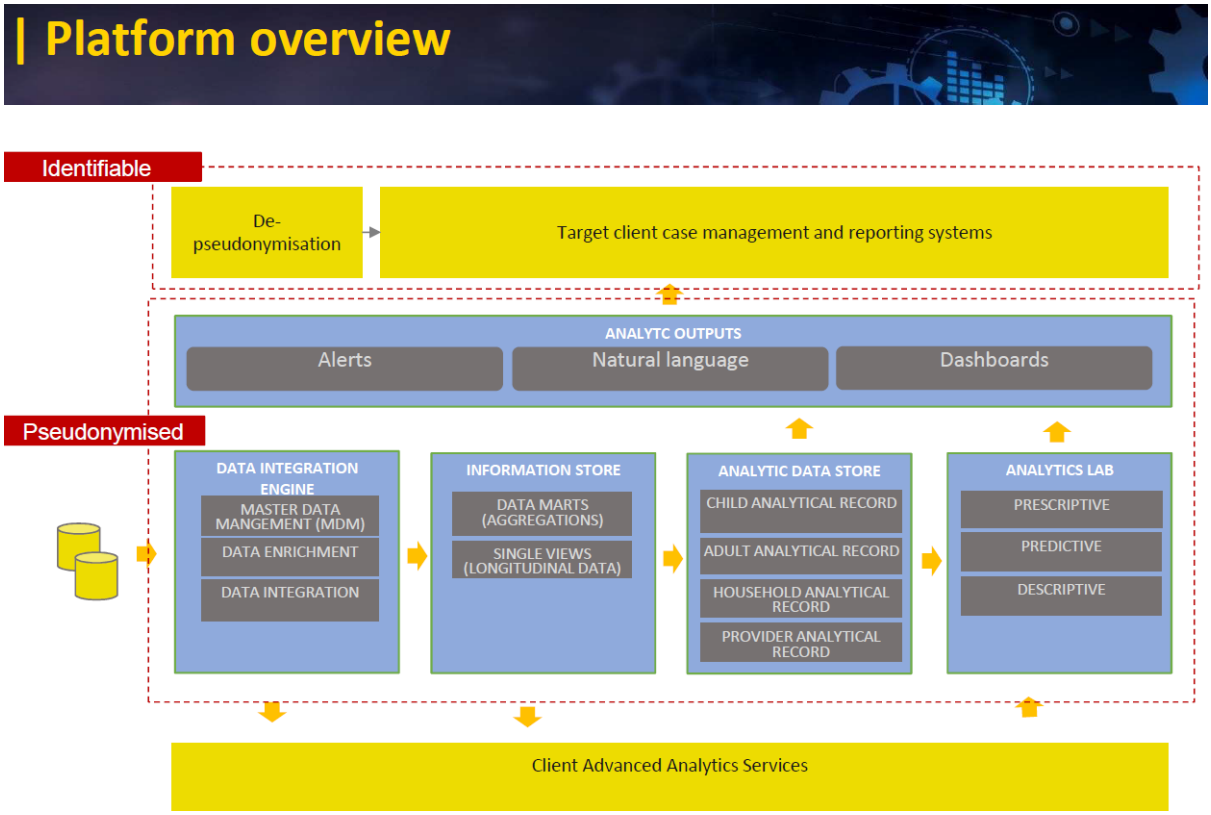
> it's educating people in the business about data collection and how bad data, especially if we're moving towards sharing data and moving towards a data lake, how bad data somewhere can start bounding around if it's not addressed and lead to more problems. So it might not matter too much to a particular service if they've got the address slightly wrong but then think about the wider implications of that. (project manager)

Another challenge mentioned in interviews was the initial hesitancy about sharing data across services, either for legal protection reasons or 'just cultural…and being concerned it would be used in a different way or in a way that is not intended for.' (business analyst) However, it was noted that this challenge is less prominent now than it was when the system was first implemented. In one interview, the point was made that the reluctance to share information is no longer appropriate:

> my personal view is that this issue, and particularly about vulnerable people dying in some cases because information isn't being shared, in some places has been going on far too long and the risk aversion around the sharing is not proportionate, and that actually it's been proven in case after case that it is ethical and proportionate and legal that a risk-based approach needs to be taken, not a sort of yes/no legalistic approach. (former employee)

In terms of the workings of the algorithm and its ability to match records, it was noted that it can go wrong in two ways: 'You can either not match people together who are the same person or you can start matching people incorrectly who are different people, and if you swing the algorithm too far one way, you get one problem and too far the other way, you get the other problem and it's what the risk is, what's the greater risk. If you join people incorrectly together, is that more of a risk than records being skewed all over the place?' (project manager)

Hackney's Early Help Profiling System



Source: https://www.ukauthority.com/events/event-hub-ukauthority-data4good-2018/

Summary

Hackney County Council makes use of data analytics in a number of ways ranging from population level health analytics, fraud detection, to child welfare. This summary focuses on uses of data analytics in Hackney for child welfare. In child welfare, Hackney County Council is working with Ernst & Young (EY) and Xantura on the use of a system to identify children at risk of maltreatment and families who need additional support. The system is called the Early Help Profiling System. It has been funded by EY and London Councils. The system is being trialed and alerts have already led to early interventions.[64]

Implementation

Xantura developed a tool they refer to as a Fusion platform "to better help services and reduce financial pressures in several business domains, including children's services, adult

[64] Stevenson, Luke (2018) 'Artificial intelligence: how a council seeks to predict support needs for children and families', *CommunityCare*, March 1, available: http://www.communitycare.co.uk/2018/03/01/artificial-intelligence-council-seeks-predict-support-needs-children-families/

social care and health, housing / homelessness and community safety'.[65] Different local authorities in the UK are using Xantura systems to address different areas. In Hackney, the the Early Help Profiling System (EHPS) is being used to identify children at risk of neglect or abuse. The system uses 'a predictive risk model which brings together data from multiple agencies to identify children who are most at risk of neglect or abuse'. The system is designed to provide social workers with monthly risk profiles that amalgamate information about families identified as most in need of early intervention.[66]

> 'The project that we've been undertaking gives us an opportunity to pick out of the 54,000 plus children who live in Hackney those children who look most likely to benefit from early help service. The system is looking to identify children and families at an earlier stage before they would get to the point of seeing social workers and to direct them to different forms of support that will help them to alleviate their problems'. (local authority)[67]

The developer of the system says it has been designed not to be punitive, but to enable earlier intervention to prevent the need for statutory intervention. The system only shares data about people who are already working with an agency or professional. The alert is sent to professionals / case workers who make decisions about the kind(s) of services or interventions needed.[68] The goal, according to those involved, is not to replace professional staff but to support them by giving them the information they need to do their job better (developers).

Across interviews, presentations and documents several contextual factors are presented as leading to the development and use of the EHPS. As detailed across the case studies, there is recognition here too that local authorities are facing a funding crisis at the same time that demands for social supports are rising. EY and Xantura argue that they can help public sector agencies use advanced analytics to improve service outcomes, reduce demand for interventions and by doing so reduce costs.[69]

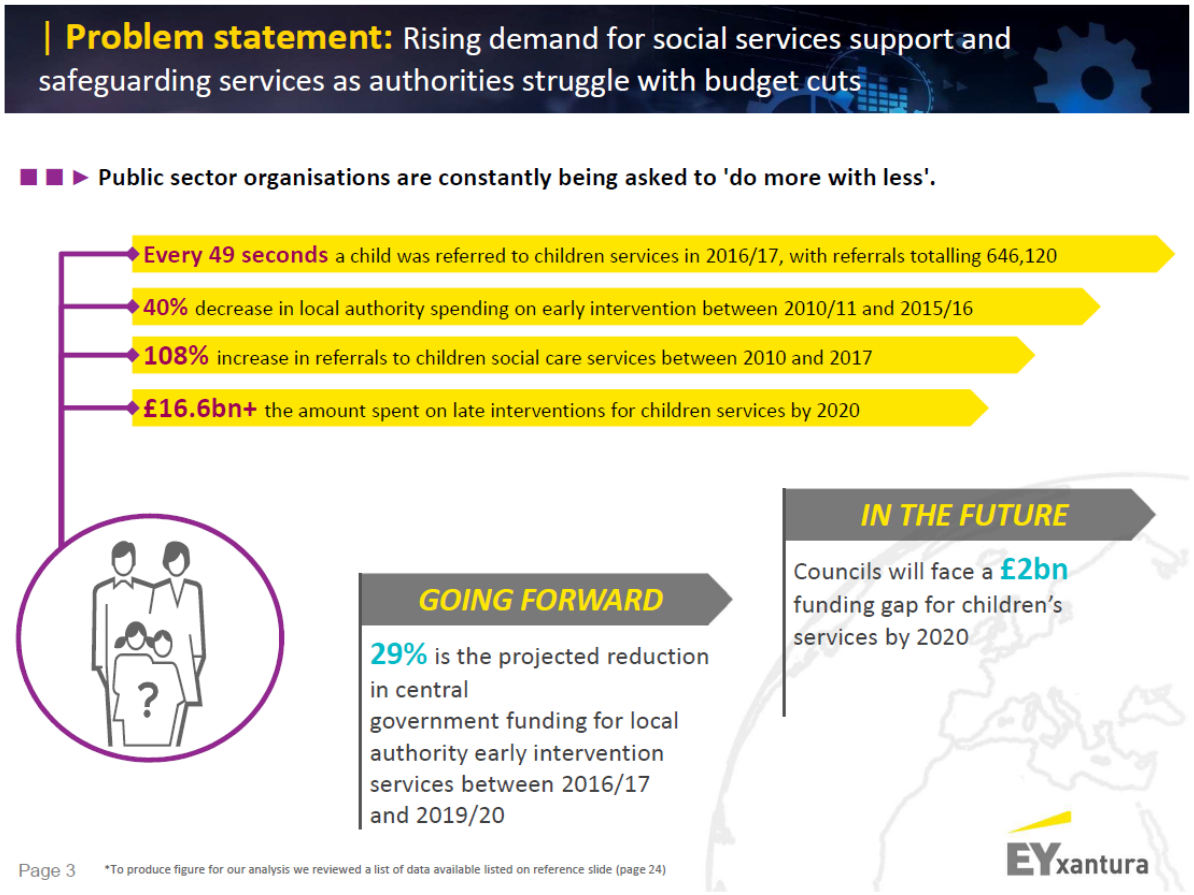
[65] Xantura (2018) Maximising the impact of early help resources, Xantura, 24 September, available: https://www.xantura.com/points-of-view/maximising-imact-early-healp-resources.

[66] London Councils (2018) Venture Spotlight: EHPS, available: https://www.londoncouncils.gov.uk/node/31412

[67] This quote comes from a comment made by Steve Liddicott in an EY promotional video.

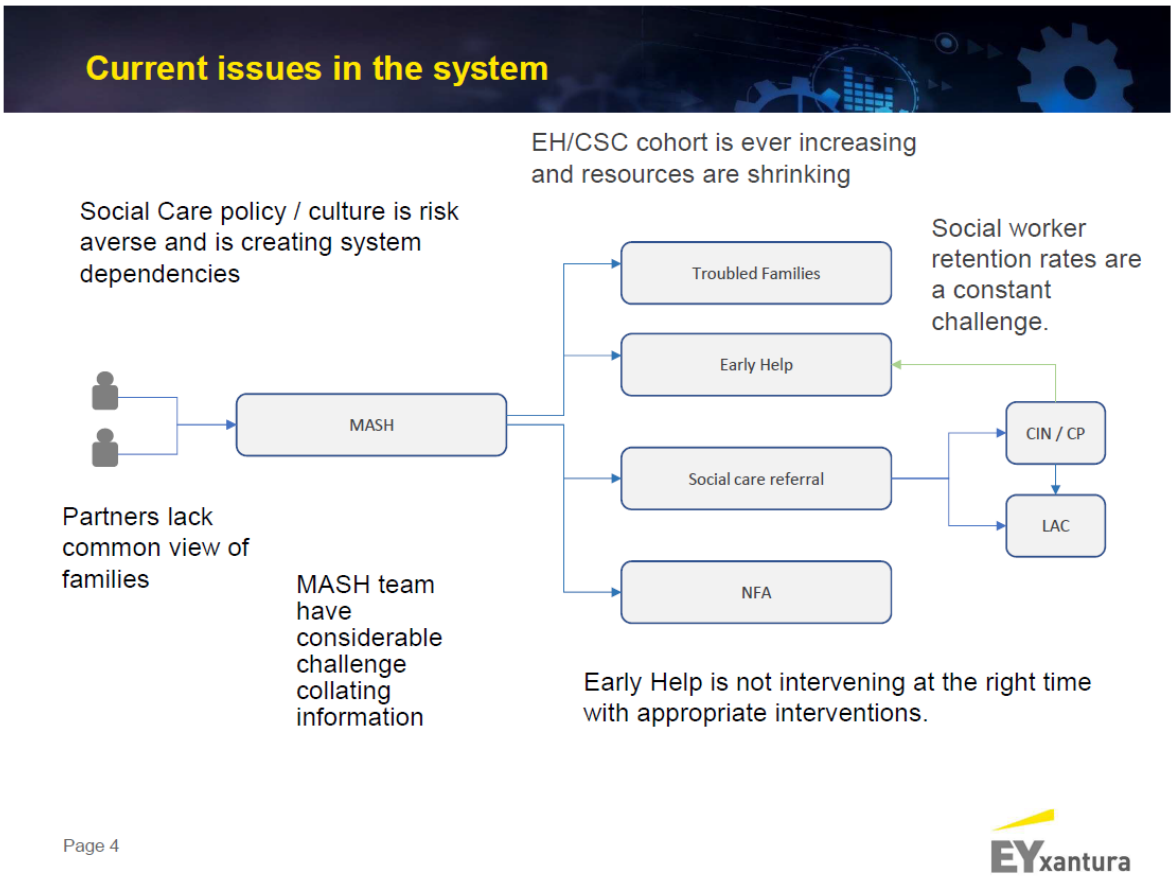
[68] Xantura (2018) Maximising the impact of early help resources, Xantura, 24 September, available: https://www.xantura.com/points-of-view/maximising-imact-early-healp-resources.

[69] Xantura (2018) Maximising the impact of early help resources, Xantura, 24 September, available: https://www.xantura.com/points-of-view/maximising-imact-early-healp-resources.



Source: Celia, Hannah (2018) Building capacity through data and analytics to improve life outcome, UK Authority, Data4Good 2018, 16 October, available: https://www.ukauthority.com/events/event-hub-ukauthority-data4good-2018/

The platform generally has also been described as addressing some problems in the social care system, particularly the challenges of sharing information about families.



Source: Celia, Hannah (2018) Building capacity through data and analytics to improve life outcome, UK Authority, Data4Good 2018, 16 October, available: https://www.ukauthority.com/events/event-hub-ukauthority-data4good-2018/

In an interview developers referenced scandals such as those surrounding the cases of Baby P, Victoria Climbié and Fiona Pilkington as raising concerns about the failure of agencies to share and act on information. These cases, say a developer, point to the need to balance privacy rights with the rights of vulnerable individuals. The system has been developed to do this by using pseudonymised data, but making data identifiable to the professionals involved when an alert is generated by the model indicating a high risk threshold has been passed (developer).

Model

The predictive models developed for local authorities, as with EHPS in Hackney, are based on local data. The idea is to build predictive analytics around outcome indicators and given the differences in data by location to ensure that local data is used to develop the modeling. They modelled nine months behind an event:

> We have spent the last 18 months building a language generation tool which takes all of the analytics, all the stats and everything else that we generate and produces a research report. The research report doesn't talk about a score of high, medium or low, it talks about the cause for concern based on the rules that our clients have defined. The outputs if the process say; this is what's happened, this is the reason that we're showing you the data, this is the family composition, this is the attendance levels, this is the chronology of events that have happened for the family, the service interventions that have gone in. And then the statistical analysis is saying, where does the family sit with respect to exclusions and is it getting worse, is it getting better, are siblings affecting younger siblings in the household? We're still tweaking and tuning, so it's not finished yet, but the idea is that we're trying to augment the process so when the social worker looks at that case, they can very quickly do, in a more accelerated fashion, the job they would normally do; not change the job. (developer)

The developers argue that the system will be able to more quickly look at and assess historical data than a case worker would be able to do. They also note that the system has been designed with the help of those working in this area.

To date there is little information available about how risk is calculated, in particular about how variables are weighted. The variables used for uses of predictive analytics in child welfare as tested in other countries has come under criticism because of the way bias can be embedded.[70]

We submitted a Freedom of Information Request to the London Borough of Hackney requesting details about multiple systems and were told that responding to our request would exceed the hourly limit councils are obliged to meet. We have submitted a narrower request and are waiting on that response. Another FOI request asked specifically for details about Hackney's use of the Children's Safeguarding Profiling System.[71] In this request it was

[70] See for example: Keddell, Emily. (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. Also: Gillingham, Philip and Graham, Timothy. (2017). Big data in social welfare: the development of a critical perspective on social work's latest "electronic turn." *Australian Social Work*, 70:2, 135-147.

[71] Knuutila, A. and Hackney Borough Council 2018. FOI request: Documents relating to the Children's Safeguarding Profiling System. Available at: https://www.whatdotheyknow.com/request/documents_relating_to_the_childr [Accessed: 16 October 2018].

stated that the Council could not reveal system details including manuals or data sharing agreements because it would damage Xantura's commercial interests:

> 'Xantura and London Borough of Hackney are working together to develop the system as development partners, but Xantura anticipates operating on a commercial basis. We believe that to reveal detailed workings of the system would be damaging to their commercial interests and, while the project is in pilot phase, of limited public use. We therefore believe that the public interest in seeing any operating manuals is outweighed by Xantura's commercial interests and exempt this part of the request under Section 43 of the Freedom of Information Act (Information Management Team, LBO).[72]

Documents, publications and promotional material note that multi-agency sources of data are used. Datasets listed include school attendance, exclusion data, housing association repairs, arrears data, police records on anti-social behaviour and domestic violence, names, addresses, dates of births, unique pupil numbers, children and adult social care, housing, debt, council tax, housing benefits and substance abuse data.[73]

[72] Ibid.

[73] McIntyre, N. and Pegg, D. 2018. Councils use 377,000 people's data in efforts to predict child abuse. *The Guardian* 16 September. Available at: https://www.theguardian.com/society/2018/sep/16/councils-use-377000-peoples-data-in-efforts-to-predict-child-abuse [Accessed: 17 October 2018]. Also: Hackney Council, Information Management Team [no date]. Privacy Notice - Troubled Families Programme.

Source: https://www.xantura.com/services/predictive-analytics

The illustration above is used in an EY promotional video to demonstrate the kinds of data used in the system. The data is pseudonymized before it is processed. [74] A developer noted that some of the data, such as the substance abuse data, has only been made available because they are using pseudonymized data (developer):

> We have names and address repositories with time series associated with them so we can see what's happening to households all the time, but this doesn't include any sensitive data. We then have a pseudonymised data repository, which is all the sensitive data matched, including substance misuse data, which clients can use for research purposes - the key point is that this is not identifiable data. The data sharing protocol rules … are set up as the key controls that mean actually I can move now from using the data purely from an analytical perspective, understanding general trends and doing analysis and building the models, to a real-world scenario, where a contact / referral is made into the front door in the MASH (multi-agency safeguarding hub), depending on reason for the contact, again defined in discussion with clients, we can run the risk model. So the risk model doesn't run all the time, the whole alerting process is only running when a certain set of circumstances are occurring. So it's a very controlled release of data from that pseudonymised repository of data. (developer)

A London Ventures summary notes that 'timely data from the vulnerable families and data from others with whom they are in contact, using mobile phone and web technology' is also being used. No further details about how mobile phone and web data are used is available.[75]

When describing the system a developer noted that particular scenarios that have been and are being developed will lead to the risk model running and potentially an alert being sent to a caseworker.

> In the children's model, say we've got an exclusion – if an exclusion is for a child just misbehaving in school and gets sent home for half a day, then they wouldn't have met the criteria and the risk model wouldn't run. If they've got a pattern of exclusion behaviour and actually it's accelerating and they punch a teacher, and this is combined with wider risk factors, for example recent youth offending or ASB activity, then the risk model would run. (developer)

[74] Stevenson, L. 2018. Artificial intelligence: how a council seeks to predict support needs for children and families. *Community Care* 1 March. Available at: https://www.communitycare.co.uk/2018/03/01/artificial-intelligence-council-seeks-predict-support-needs-children-families/ [Accessed: 16 October 2018].

[75] London Councils [no date][a]. Keeping children safer by using predictive analytics in social care risk management. Available at: https://www.londoncouncils.gov.uk/our-key-themes/our-projects/london-ventures/current-projects/childrens-safeguarding [Accessed: 15 October 2018].

In terms of accuracy it has been noted that 'Over 80% of households in Hackney that have been identified most at risk by the model are at risk'.[76] However, in order to judge the accuracy of the model it is also important to know how many people were wrongly identified as high risk when they are not. It is not clear how the implications for those wrongly identified as high risk are considered or what opportunity people have to challenge or remove from systems a high risk assessment that is wrong.

In terms of consent, those subjected to the system are not being informed that their data is being used. The argument is that releasing details may prejudice potential interventions and compromise the commercial interests of the company involved, Xantura. It's also noted in the Privacy Impact Assessment that there is no option to opt-out of having data included.[77] The public more generally has not been consulted about the development and use of this system.

Deployment and Uses

The ideas behind the development of the system were described in 2015 by Hamza Yusuf, then head of finance at the London Borough of Hackney:

> 'The premise is that academic research and the government's Troubled Families Programme identify a number of risk factors, both distal and proximal, related to child maltreatment. These can include benefit receipt, a history of offending, poor educational attendance or issues with parental capacity. The model we aim to pilot will identify these risk factors in a given family and, where they present collectively, an alert will be sent to a children's services practitioner or our multi-agency safeguarding hub to investigate further. The model will help us to identify children who are at risk of maltreatment and target our interventions more intelligently, to prevent escalation into statutory social care'.[78]

After a family or person has been identified as benefiting from early help the idea is that the case worker uses information in the report to identify the range of issues affecting the family to provide more support (developers).

> 'What quite often we see is there's a very definite cause and effect. Someone is struggling in school and so there might be an educational officer that might be able to support them, but if that educational officer doesn't know that there's been a challenge with mum maybe and mum's struggling financially or she's just lost her job, what we can

[76] London Councils and Ernst & Young LLP 2017. London Ventures. Guide to London Ventures. Available at: https://webforms.ey.com/Publication/vwLUAssets/EY-london-ventures-guide-april-2017/$FILE/EY-london-ventures-guide-april-2017.pdf [Accessed: 15 October 2018].

[77] London Borough of Hackney. (2017). FOI response to requst by Mr. Knuutila, 17 April. URL: https://www.whatdotheyknow.com/request/documents_relating_to_the_childr#incoming-1143765 [10 September 2018].

[78] Yusuf, Hamza (2015) Calculating a positive future: using big data to manage demand and make savings, Public Finance, Sept. 11, available: https://www.publicfinance.co.uk/opinion/2015/09/calculating-positive-future-using-big-data-manage-demand-and-make-savings

do is highlight to the lead professional that there is also a broader stress in the household' (developer).

Xantura and the London Borough of Hackney are working together to develop the system, but Xantura expects to operate commercially.[79] The EHPS is being promoted as a system that can help councils save significant amounts of money by preventing family situations from escalating to the point where children are taken into care.[80]

It has been suggested that the screening of children and families could become fully automated and that the data could be used to generate snapshots of families that could be used in referrals going forward. A pilot project with general practitioners to assist them with making referrals has also been noted.[81]

Auditing and Safeguards

A Privacy Notice indicates that data is encrypted and held in a secure facility. It is also noted that 'only those council employees who currently use similar data will have access to the information supplied by Xantura (and) therefore current council vetting and tracking procedures will apply. It is stated that access to identifiable data will be through 'a secure application with user access controls' to ensure that only authorized people can access the data. It is also noted that 'identifiable personal data will only be made available if certain business rules (based on an assessment of risk and vulnerability) have been met.[82]

Access to data and sharing of data is said to happen through data sharing protocols and that every time data access occurs it is logged. Developers noted they are building a tool for council information governance so that reports can be generated that indicate everyone who has used the data and for what purpose (developer).

Recent media coverage indicated that the Information Commissioner's Office is looking into the use of children's data in this case. There has been no public report made.[83]

Evaluations of the system are said to be iterative and ongoing, with the 'performance of the predictive aspects of the system being tested statistically and piloted in a live setting'.[84] In an interview it was noted that at present there was not a capability at this time to evaluate the impact the system was having on service users.

---

[79] Information Management Team, ICT Services, London Borough of Hackney 2018. FOI request by Aleksi Knuutila - response 17 April 2018. Available at: https://docs.google.com/document/d/1SyZfOCu5PW8X1eRdfPTw0Bf4X4R0O7hyd2pBMvPJlHM/edit?usp=drive_web&ouid=108885959520275379582&usp=embed_facebook [Accessed: 15 October 2018].

[80] Yusuf 2015 and Graham 2017.

[81] Graham 2017 and Stevenson 2018.

[82] Hackney Council, Information Management Team [no date]. Privacy Notice - Troubled Families Programme.

[83] Adams, Joel (2018) Councils using 'hundreds of thousands

[84] Hackney Council, Information Management Team [no date]. Privacy Notice - Troubled Families Programme.

Developers say they are able to validate their models, but that tracking the impact of their alerts and related changes are challenging because councils do not have the baselines needed for them to measure and track impact (developer):

> Operationally the benefits case isn't just about cost avoidance, for example, the natural language capability benefits case is about whether we can improve the quality and reduce the time it takes for a MASH worker to go from contact through to assessment close, because by gathering all this data and presenting it to the social worker it takes less time to compile this research. In order to track these benefits we're now creating reports that allow us to track contact to close on average by source. (developer)

At the moment, there is no way of measuring if and how this new system is affecting those using these services, to gain their perspective on how effective the early interventions have been.

Developers note that the system provides a means to monitor for bias:

> [I]f we look at the child protection caseload, we can see, for example, what the age distribution of the children are, I can see what the ethnicity distribution for this client is, I can see what all those different characteristics are, what the deprivation is. For each of those distributions, I can say what does the model do, is it a different distribution to the distribution in the client's caseload? If it's significantly different and the model is skewing oddly, why has it got a different bias to what's naturally in the data? This could be because there is bias the client's existing system or that the model is biased? So for the first time, I think, we can actually start looking at bias in the system, which is actually quite a powerful thing to be enabling. (developer)

This type of comparison will not address the extent to which other biases may enter a system, as argued by people like Virginia Eubanks, through assumptions about what is normal and what a family should look and act like.

Challenges

The following charts describes the lessons learnt during the implementation of our models and solutions to mitigate the issues.

Issue	Lesson learnt	Approach to mitigate this issue
Adequate resources are not made available	Appropriate staffing resources will need to be agreed at design phase from all relevant steering groups	Identifying specific tasks and responsible parties will help to support mobilisation and implementation, project progress and resourcing arrangements will be reviewed on a regular basis
Partners are not willing to share data for data protection reasons	Limitations of data accessibility limits the capabilities of our solutions	We will work with local data guardians in partner organisations to establish a staged, proportional data disclosure model that operates on a 'need to know' basis – and has been used elsewhere. i.e. The identification of families that are already known to a service (error identifications) decline with more data accessibility
Partners are not willing to share data for operational / cultural / technical reasons	We need to ensure all relevant stakeholders understand the impact of our solution within their service and for children and family services as a whole to achieve a more collaboratively approach	During the initial phase, we will ensure through a robust programme management approach that we will schedule meetings and workshops to gather perceptions and identify opportunities and risks so strategic and operational buy-in is achieved
Data quality issues prevent meaningful analysis	Although data quality issues do exist, our experience is that there is sufficient high quality information available to support the development of predictive models	Data cleansing routines will be applied to all data supplied to the system, additionally, all data is assessed in terms of quality before it is included in advanced analytics processes.
Professionals might consider the solution as a replacement of their professional judgement or a deletion of their jobs	To reduce job security anxieties and to enable professionals to use the solution as an enabler of their capabilities, it is important to work together to manage resistance to change	A desk based pilot will ensure the solution is implemented in a way that augments and supports professionals to highlighting potential risks and providing wider information in a more streamlined manner
Forecast savings are not realised as services fail to incorporate the tool and exploit its potential benefit	We need to have a clear and auditable benefit target that is signed off by the steering group at design phase and there needs to be is incentives for both parties to achieve this	Upfront, we will agree with you the detailed project plan which will include activities by EYXantura and the local authority of what is required to ensure the agreed benefits and savings are delivered. An outline of the business case will also be drawn up during the design phase

Source: https://www.ukauthority.com/events/event-hub-ukauthority-data4good-2018/

One of the challenges identified by developers interviewed is that others will not differentiate between the system they have developed and other uses of predictive analytics in child welfare in other countries.

Xantura identified a range of challenges have been identified through their efforts to make use of predictive analytics in partnership with local authorities during a local government event.

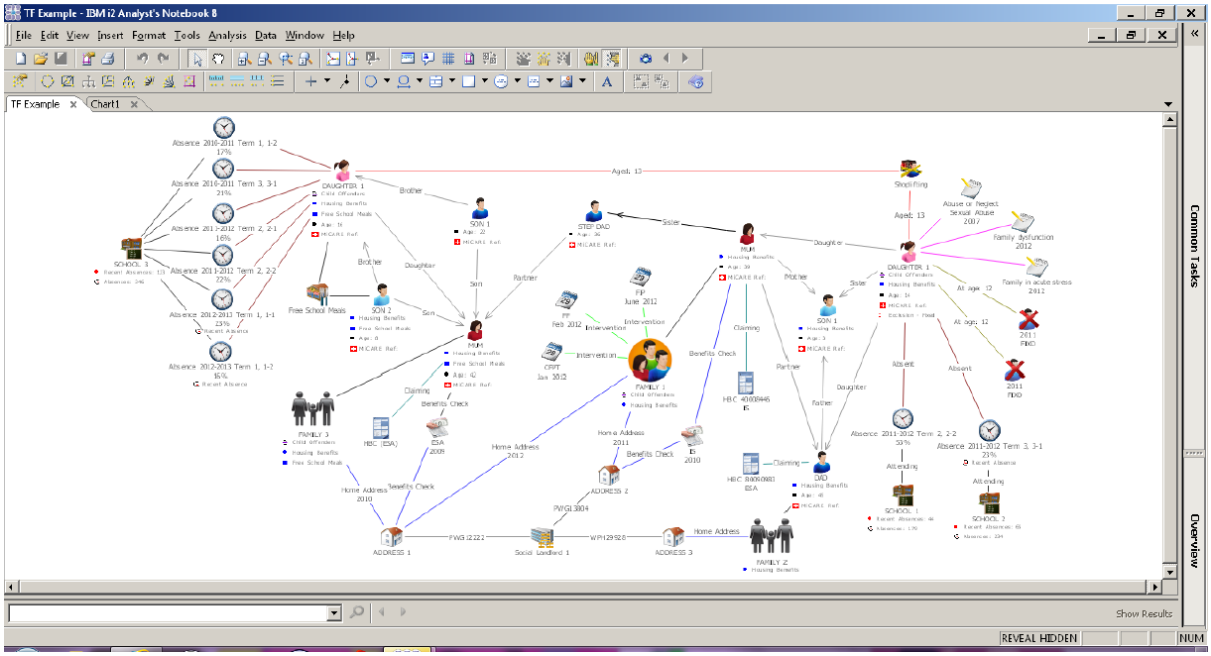
These are listed as inadequate resources, unwillingness to share data, data quality, anxiety about job replacement or loss, and a failure to realise the savings promised. Each of these warrant more public debate.

Concerns about privacy and the public's attitudes to how systems like these might invade privacy are an ongoing concern.

There is interest in expanding the uses of predictive analytics in child welfare and social services globally. EY is said to have a global team looking at this.

Manchester's Research & Intelligence Database

8. Finding the Connection



Source: https://www.local.gov.uk/sites/default/files/documents/childrens-social-care-and-b32.pdf

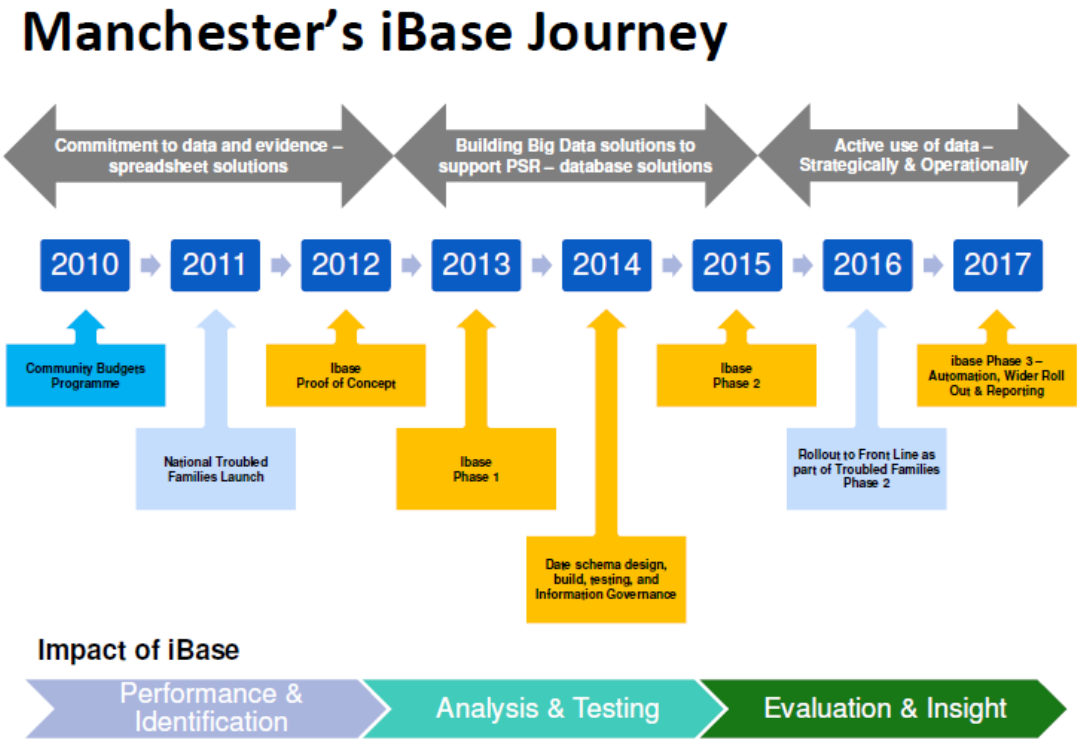
Summary

Manchester City Council is using a system it calls the Manchester Research & Intelligence Database to identify Troubled Families, families in need of support, to enable caseworkers to access more information more quickly than previously and to enable the collection and analysis of data to assess services and impact. The Council developed the system by buying an IBM product called iBase. The approach to using the system has been internally developed. It is noted that the aim is to empower 'lead professionals to make the best use of data they are legally able to see, and is not about replacing decision making or interpretation with system algorithms or decisions'.[85] The stated aim is to enable a more 'holistic' understanding of people, needs and services. Future goals include developing decision making tools, building performance reporting tools, building threshold and alerting tools and rolling the system out to more users, as well as connecting more data.[86]

[85] Manchester City Council (2018) Freedom of Information Response, CHS/AWZBQX, 17 April, available: https://www.data-scores.org/b7137a1bd1c89100e5812a78b67e83ea8bc5ac005d6a7029e0bbb0d7b462586d

[86] Holme, P.; McNichol, L., Manchester City Council 2017. Ibase in Manchester, available: https://slideplayer.com/slide/13557071/

Implementation



Source: https://slideplayer.com/slide/13557071/

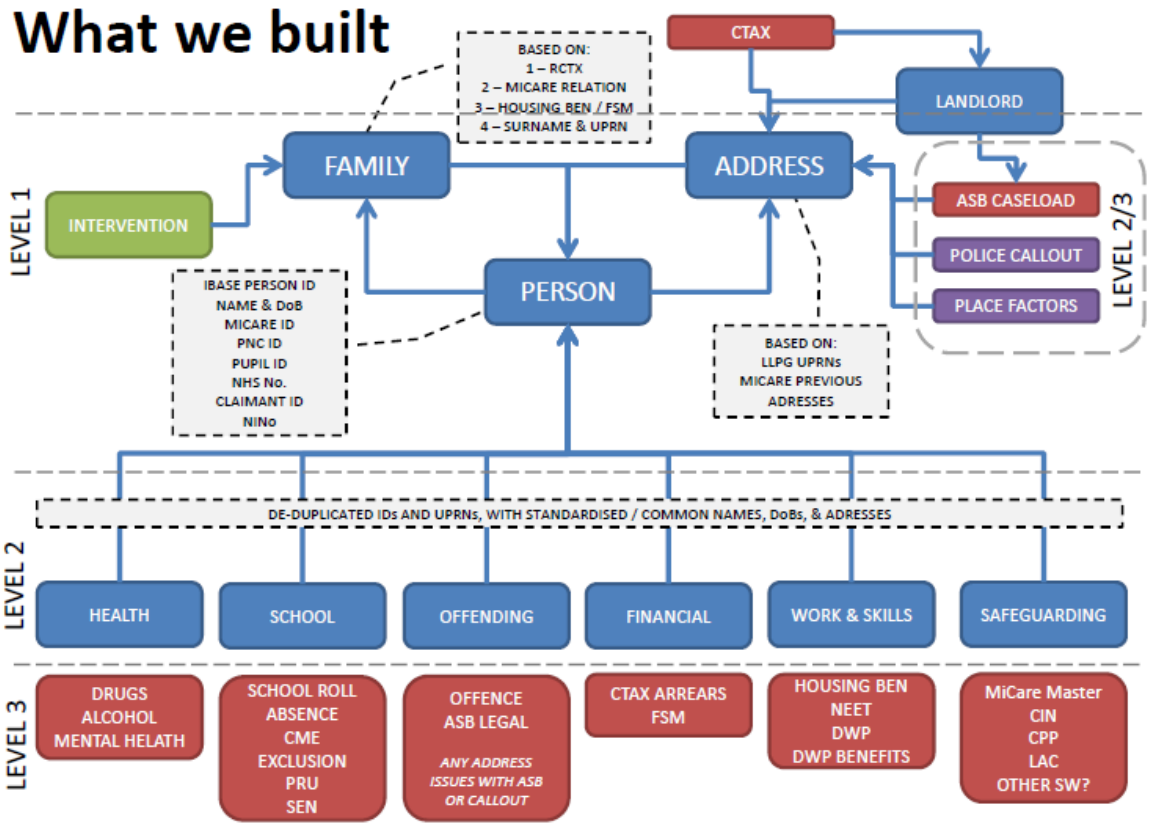
Manchester's use of data analytics was motivated, in part, by the need to identify families that met the criteria for the Troubled Families programme. In Manchester, the experience was that a data system that enabled the ability to understand complex and multiple factors as well as networks was required.

The data warehouse that Manchester created, as of 2016, integrated 16 datasets. Caseworkers are able to access this data going back five years.[87] The aim was to employ a system that links individuals to families and households, to organisations like schools and that also reveals the links between families and that tracks interventions, service use and outcomes. Other goals were to ensure visualization as a means of assisting case workers, and to also employ a system that could be used to make decisions about service reform and to enable finding efficiencies and savings.[88]

[87] Symons, T. 2016a. Wise Council. Insights from the cutting edge of data-driven local government. Nesta and Local Government Association, available: https://media.nesta.org.uk/documents/wise_council.pdf

[88] Henry, Sarah (2016) A families database – from an individual to a holistic view, Local Datavores Workshop, available: https://www.youtube.com/watch?reload=9&v=M128IZ1OltQ

In public presentations it has been noted that consent is sought from those whose data is used in the system because it is viewed as good practice to tell people their data is being shared, but that there are also occasions when people's data is shared without consent.[89] The Privacy Impact Assessment is no longer available online so more details about how consent is sought and from whom is not available.



Source: https://slideplayer.com/slide/13557071/

Model

The iBase system, IBM i2 iBase IntelliShare was purchased in 2012 as part of Manchester City Council's Troubled Families initiative. It was determined that this was the best product to enable the data matching and visualisation needed to work on the Troubled Families programme.[90] This was purchased as an 'off-the-shelf' product and there was no collaboration with IBM to produce the product. There is no development or data sharing agreement with IBM.

The datasets used are detailed in the model above and updated on a regular basis. These are imported from internal and external systems and are split into the following categories:

[89] Henry, Sarah (2016) A families database – from an individual to a holistic view, Local Datavores Workshop, available: https://www.youtube.com/watch?reload=9&v=M128IZ1OltQ

[90] Waterhouse, K., Manchester City Council 2018. FOI Reference TREAWZBQX Internal Review Aug 2018

address data, health data, school data, offending data, financial data, work & skills data, and social care data:

> The majority of the data-sets contain details of individual people and details of the event linked to those individuals, ie offences, school absences, etc. People from the different data sets are then created in the Research & Intelligence database using their first name, surname & date of birth as the unique identifier linking any events and addresses relating to this person…
>
> The creation of a family entity is created once a family has been referred into a service and a lead professional has completed a whole family genogram, the lead professional then adds a family number to each individual from the family which links the individuals together in the system.[91]

The iBase system is used to match data so caseworkers can view and access multiple sources of data at the same time easily. It is also used to identify connections and relationships between individuals and families.

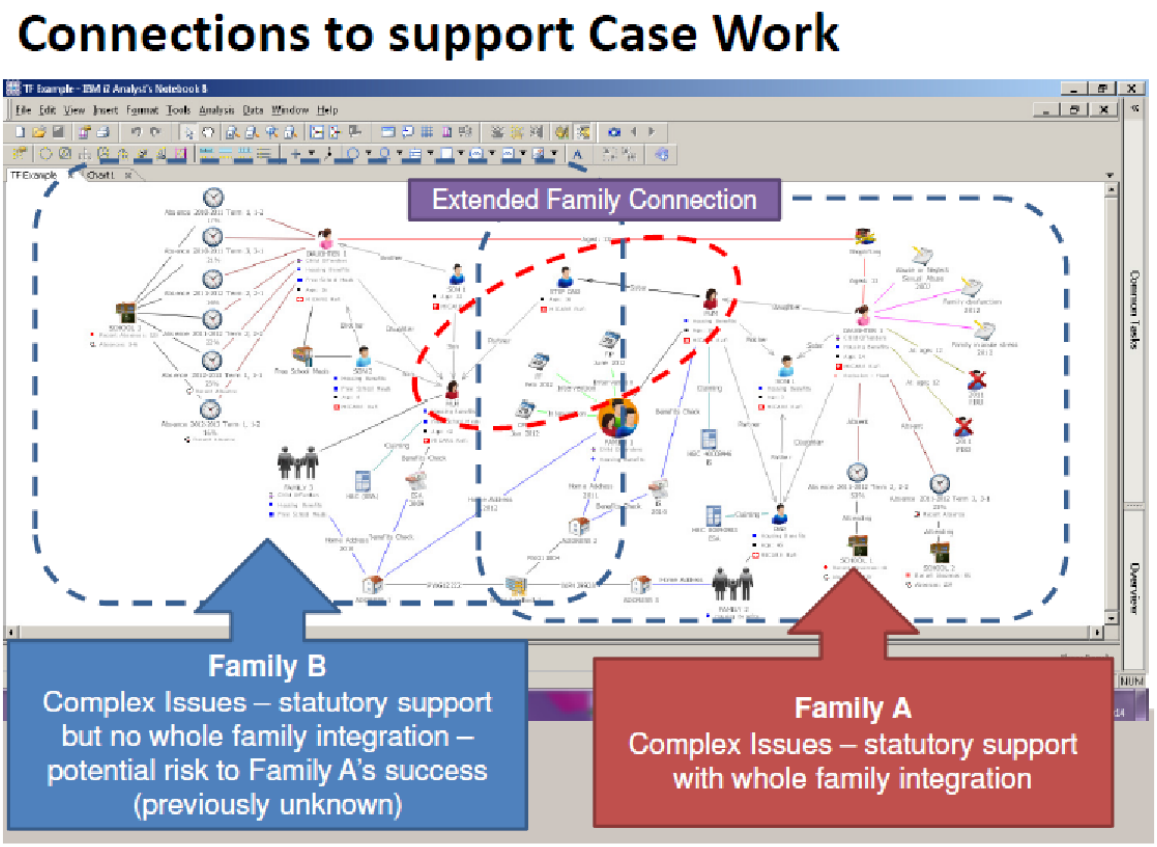
Data used for analytics by researchers and analysts is extracted from the system and anonymized.

Predictive modelling is used to determine where help is needed, where it will succeed or not and where families can help themselves. In terms of the techniques being used, decision trees are used to try and find causality links and predict when a child in need status will be flagged in a system. Cluster analysis is used to identify distributions of needs or characteristics of families in order to understand how to shape a programme or assign support. Regression analysis is used to identify factors that are most important and predictors of future events. Spatial analysis is used to plan services and understand the relationship of spatial issues to other policies.[92]

---

[91] Manchester City Council (2018) Freedom of Information Response, CHS/AWZBQX, 17 April, available: https://www.data-scores.org/b7137a1bd1c89100e5812a78b67e83ea8bc5ac005d6a7029e0bbb0d7b462586d

[92] See: Henry, Sarah (2016) Local Datavores Workshop, LGA, July 13, available: https://www.local.gov.uk/sites/default/files/documents/childrens-social-care-and-b32.pdf; Symons, T. 2016a. Wise Council. Insights from the cutting edge of data-driven local government. Nesta and Local Government Association, available: https://media.nesta.org.uk/documents/wise_council.pdf.

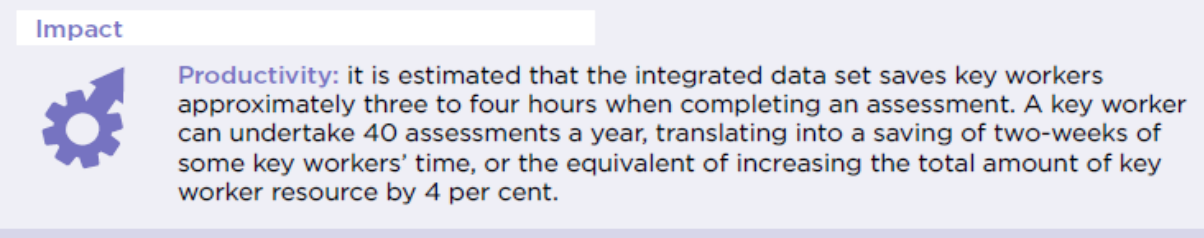


One of the aims of the system is to give caseworkers better quality information. About 200 frontline users are said to have access to the intellishare product, the webportal that users use to access iBase in order to query the database about their cases.[93] It is described as giving caseworkers, 'in just a few clicks' a comprehensive view of a family, including interactions with other agencies, as well as social and family networks. When a case is presented to a worker, the worker looks in the iBase system to understand the activities around the family and then uses this information to direct their follow-up questions and searches.[94]

Another stated aim is to save keyworker time. As detailed in Nesta document:

[93] Holme, P., Manchester City Council 2017. Paul Holme Research and Intelligence Manager - Manchester City Council iBase in Manchester. Available at: https://vimeo.com/214846064 [Accessed: 13 September 2018].

[94] Holme, P., Manchester City Council 2017. Paul Holme Research and Intelligence Manager - Manchester City Council iBase in Manchester. Available at: https://vimeo.com/214846064 [Accessed: 13 September 2018].



Source: https://media.nesta.org.uk/documents/wise_council.pdf

We were unable to secure an interview with frontline workers to determine the accuracy of this estimate. More work generally is needed to better understand how frontline workers are engaging with this technology.

Performance management is another way that the system is used.

> '[T]he software also makes it easier to monitor and check social workers' cases for managers, and provides a useful set of checks at the point at which a case is closed'.[95]

It is also noted that performance data is combined with cost data to inform the 'investment case' for early intervention programmes.

> For instance the system was used to inform an investment case for Multi-Systemic Therapy (MST) by accurately identifying likely cohort size. MST involves teams of four staff, and requires a minimum cohort size. Correctly identifying the likely demand for services prevents wasted expenditure.
>
> The system provides evidence about the effectiveness of different interventions such as FIP (Family Intervention Programme) or Early Help, which can be used to decide on re- and de-commissioning of services or how to reshape service provision.[96]

The data is also used for analytics. Data is extracted from the iBase system, anonymized, and then analysed by researchers and analysts through the use of statistical packages such as Excel, SPSS, R and MySQL.[97]

Auditing and Safeguards

The approach to using the iBase system was developed internally. The Council notes that: 'no personal data has been shared with agencies / collaborators for the purposes of system design or analytical reporting'.[98]

---

[95] Symons, T. 2016a. Wise Council. Insights from the cutting edge of data-driven local government. Nesta and Local Government Association, available: https://media.nesta.org.uk/documents/wise_council.pdf

[96] Symons, T. 2016a. Wise Council. Insights from the cutting edge of data-driven local government. Nesta and Local Government Association, available: https://media.nesta.org.uk/documents/wise_council.pdf

[97] Manchester City Council (2018) Freedom of Information Response, CHS/AWZBQX, 17 April, available: https://www.data-scores.org/b7137a1bd1c89100e5812a78b67e83ea8bc5ac005d6a7029e0bbb0d7b462586d

[98] Manchester City Council (2018) Freedom of Information Response, CHS/AWZBQX, 17 April, available: https://www.data-scores.org/b7137a1bd1c89100e5812a78b67e83ea8bc5ac005d6a7029e0bbb0d7b462586d

The system has been subject to a Privacy Impact Assessment, although we were unable to access this online at the time of writing.[99]

The system tracks use and staff are told that unauthorized use can lead to being fired.[100]

Manchester City Council said that it sought legal guidance when developing its approach to using its integrated data system.

> Through the new system, key workers do not have to ask to share data, they are entitled to see data in line with the responsibilities and duties of their job. The decisions built into the system were based on meticulous work to decide the legality of sharing data in very specific, defined instances. This includes sharing school attendance information, or reports of police call-outs, where there is a safeguarding concern about the child. Advice from a barrister was sought on these specific instances, which then informed the development of a Privacy Impact Assesssment in collaboration with partner agencies. Manchester wanted to reduce individual decision-making about data sharing because different interpretations led to inconsistencies in which data was being shared.[101]

It is noted that a small amount of data cleansing takes place before data is imported into the system to ensure that names and date of birth are in a consistent format. 'This involves reviewing unique person IDs in the source systems to ensure that duplicates are removed and the names & DOBs matched are consistent'.[102]

Challenges

It has been noted that half of the work in developing the system related to working out the informational governance issues. Negotiating technical issues were found to not be as difficult as expected, although poor data quality was mentioned as something that posed a particular challenge.[103] Other related challenges included developing and agreeing on rules and processes as well as the 'data architecture'.[104]

---

[99] This impact assessment can be found here: http://www.manchester.gov.uk/info/200031/data_protection_and_freedom_of_information/6947/research_and_intelligence_database

[100] Henry, S. (2016) A families database – from an individual to a holistic view, Local Datavores Workshop, available: https://www.youtube.com/watch?reload=9&v=M128IZ1OltQ

[101] Symons, T. 2016a. Wise Council. Insights from the cutting edge of data-driven local government. Nesta and Local Government Association, available: https://media.nesta.org.uk/documents/wise_council.pdf

[102] Manchester City Council (2018) Freedom of Information Response, CHS/AWZBQX, 17 April, available: https://www.data-scores.org/b7137a1bd1c89100e5812a78b67e83ea8bc5ac005d6a7029e0bbb0d7b462586d

[103] Symons, T. 2016a. Wise Council. Insights from the cutting edge of data-driven local government. Nesta and Local Government Association, available: https://media.nesta.org.uk/documents/wise_council.pdf

[104] Henry, S. (2016) A families database – from an individual to a holistic view, Local Datavores Workshop, available: https://www.youtube.com/watch?reload=9&v=M128IZ1OltQ

A lack of engagement and the fact that the system was not viewed as a priority by all stakeholders has been referenced as a challenge.[105]

The documents referenced do not often identify the risks involved. Those using the system would be able to best identify these.

[105] Symons, T. 2016a. Wise Council. Insights from the cutting edge of data-driven local government. Nesta and Local Government Association, available: https://media.nesta.org.uk/documents/wise_council.pdf

Avon & Somerset Police Qlik Sense

Summary

Qlik Sense was first piloted by Avon & Somerset Police in 2016 and now has over 30 applications across teams. It serves as both a performance assessment tool and a predictive policing tool. Developed in part as a response to on-going austerity measures, the system is a form of self-service analytics software that connects internal datasets within Avon & Somerset Police, as well as some datasets from other agencies in Bristol Council, to provide integrated assessments and evaluations. The focus in this report is the predictive modeling for individual offenders and victims as well as neighbourhood mapping of crime. Built into Qlik Sense applications are offender risk scores and vulnerability risk scores along with a harm rating that determines an overall risk. It is intended as a 'one-click' system that provides individual offending and intelligence profiles to help 'triage' risks and threats. The system is used by frontline staff to decide on allocation of resources and pathways of managing highest risk offenders. In some instances, such as domestic abuse, it is used to decide on who to manage and to enable pre-emptive measures.

Implementation

Qlik Sense was piloted by Avon & Somerset Police in 2016 and put into use in January 2017 across different parts of the police force. In interviews it was noted that the decision to introduce Qlik Sense in Avon & Somerset Police work came in the context of on-going austerity measures, with around £80 million cuts in Avon & Somerset Constabulary, and attention towards developments in technology amongst the leadership team. One previous manager, who worked on developing the system within the police force said, 'there became an opportunity for Avon and Somerset to say, actually we don't want to keep having to do things like we used to do it because we're just not going to be able to survive. There's a tipping point in the organization and we have to do something to enable to do things differently in the world, like the modern world's doing.' (former employee) He went on to note, 'it's viewed very much as a critical enabler, strategic imperative for any…organization that's facing cuts.'

Qlik Sense noted in a press release relating to the contract with Avon & Somerset Constabulary that the analytics platform is used 'to visualize its command center operations data to gain better insight into the availability, objectives, and location of its police officers against public demand.'[106] It started as a management tool, collecting performance data in relation to staff, such as 'how many crimes they're managing, whether they're contacting victims, whether those crimes are being reviewed by sergeants.' (chief inspector) From that it 'mushroomed' and 'now It's really the data help around which everything revolves.' (chief inspector) Whilst it initially developed to visualize data in 12 apps, it was noted in interviews that Qlik Sense now has over 30 apps in use by Avon & Somerset Police for different functions and by different teams, and has around 4000 licenses issued across frontline staff. This includes the offender management app that uses predictive modeling and profiling for offenders, including level of risk, cohort and crime pattern. It was noted in an interview that there are about 250,000 offenders within the Avon & Somerset area that are given a score (coordinator).

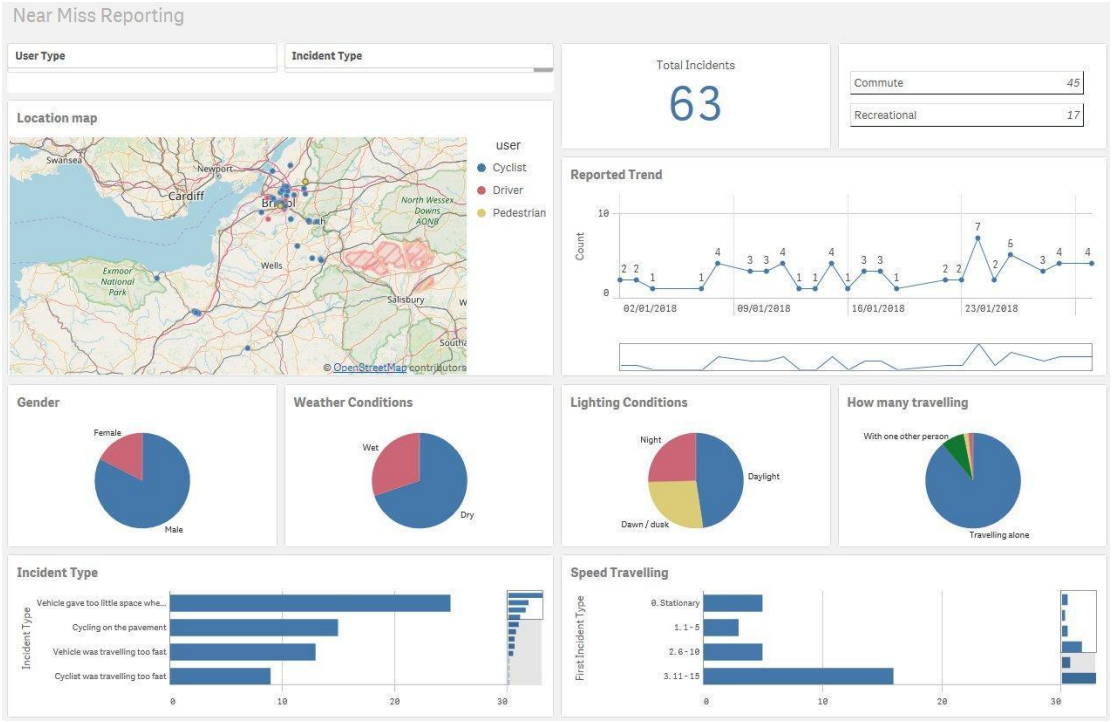
The system builds on previous predictive modeling software developed by Avon & Somerset Constabulary using IBM Predictive Analytics for tackling domestic violence and child abuse in which risk scores are produced based on 'historical crime data, along with textural and sentiment analysis combined with additional databases and open-source information, to create a statistical model that can predict an individual's behaviour and risk…allowing officers to identify potential victims before they are harmed.'[107]

Model

Qlik Sense was described in interviews as a form of 'self-service analytics software' in which different officers can access different parts of the system to locate information relevant to them. The data that informs dashboards and predictive modeling was summarized in an interview as being 'primarily internal datasets, so that is our call handling data, our crime intelligence data, missing people data, our command and control data and obviously our HR and finance and more backend datasets, operational data like airwave data so the GPS pings from officers, sourcing from user satisfaction survey data and internal staff survey data.' (manager) Currently, this might be expanded through further data sharing between agencies and the use of partner data (such as Council held data) using Home Office Transformation funding as well as information sharing with voluntary agencies (chief inspector). In one interview, it was also noted that for aspects of predictive modeling further external data is used like, 'weather models, weather forecasting stuff coming in to influence crime trend and patterns.' (former employee) (see dashboard image below)

[106] https://www.qlik.com/us/company/press-room/press-releases/0111-police-force-visualizes-incident-operations-data-fight-crime-faster-improve-public-safety

[107] https://www.computerweekly.com/news/2240231347/Avon-and-Somerset-constabulary-targets-domestic-abuse-with-predictive-alatyics

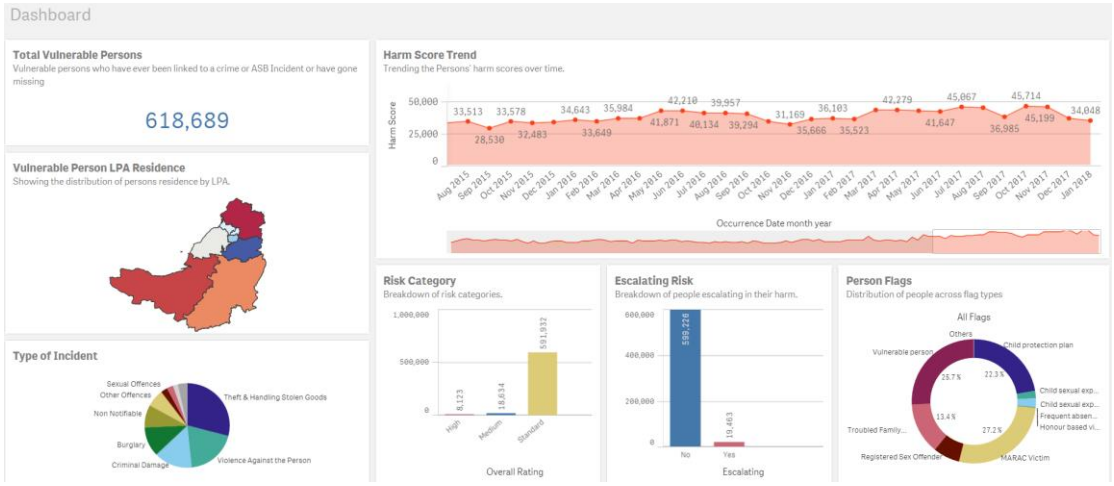


It was also noted in the same interview that ‘they do use also some social demographic information, like the Acorn type information that comes in, and that’s used by looking at areas of high crime rates, deprivation and looking to be able to support any outliers or any areas which need different sorts of interventions to help spot that.’ (former employee) However, it was noted in a different interview that Avon & Somerset police do not use demographic data such as ethnicity as part of the predictive modeling of individuals or third party modeled data, such as that provided by Mosaic or Acorn.

With the police-held data, offender risk scores and vulnerability risk scores are produced. The offender risk score will be a percentage score between 0 and 100 that identifies ‘the likelihood of offending’ in combination with the harm that an offender carries: ‘So if it’s an offender that’s previously done rapes or GBHs [Grievous Bodily Harm] or attempted murders or threats to kill, a number of things, this harm rating, combined with the likelihood of offending, allows you to determine overall risk for that offender.’ (former employee) That means that one of the variables that are weighted are type of offence: ‘it’s weighted for people who are going to commit a violence offence, [they] are scored higher, at more risk.’ (inspector) A section of the model identifies the risk of committing a serious domestic abuse offence which is based on data about previous offenders of domestic abuse: ‘we put in all data of, say, somebody that has already committed a domestic abuse offence and we put in their characteristics. So this percentage is how many of those characteristics that person fits.’ (inspector) The model also includes an escalation risk: ‘once you’re measuring risk in an automated way, you can then measure the escalation risk. So if someone’s offending behavior changes over the last week or two or even overnight, the model will then show you that and it’ll push it up the list.’ (former employee) In that way it distinguishes between risk associated with previous offences and escalating risk associated with recent or current offences: ‘So it enables you to, if you

have some that are a very similar score, it decides…which one of those is escalating. So that means that their offending behaviour is happening now.' (coordinator)

The vulnerability risk score is modeled with a similar approach to the offender risk score, providing a percentage score for the likelihood of an individual becoming a victim of crime based on data such as 'if you've been a victim of crime but also personal antisocial behaviour and missing persons incidents.' (manager) It was noted in interviews that parts of the system are linked up with Bristol's Integrated Analytical Hub (see case study in this report) 'to help, in Avon and Somerset Police's case, to determine potential long term missing people' (former employee) as one aspect of vulnerability. However, it was also mentioned that the accuracy rate for offenders is more accurate than it is for vulnerable people (coordinator).



Deployment and Uses

Qlik Sense is a suite of analytics applications of which predictive modeling is just one aspect of the system. In interviews it was noted that analytics feeds into 'everything from a strategic level' (former employee) including strategic decision making for the organization and governance of the organization and management. There are continuous analyses fed to different tasking processes; what was described as a '24/7 live cell' in which analyses go out to the organization: 'here's our top wanted offenders, that goes out to local teams, neighbourhood teams, [to those with] responsibility to manage offenders.' (former employee) It is used for crime trends daily, such as mapping burglary crime trends in relation to what there would 'normally' have been of those crimes. A key feature of the system is therefore to inform police 'in relation to demand' and to use it 'to decide which people we are going to go after out of this big list [of people who are wanted].' (coordinator)

Individual risk scores are used to alert officers to different forms of risk. Police officers are provided with some guidance as to how risk scores have been calculated, which can help them make sense of it. Including this aspect in the interface of the dashboard is important as it was noted in interviews that there is an emphasis that the model is not driving decision-making, but that professional judgement is: 'it is just a tool so we wouldn't just go down the route of saying that person's the highest score, they're the most risky people, we also look at here and now. So if there's a job that comes in and someone's, say, threatened someone with a firearm, then there's always a professional judgment that we use around the experience of people that work within our unit.' (coordinator) This framing has been key to its deployment with front-line staff. One inspector explained the nature of the initial hesitancy around the model:

> When we first started looking at it, we couldn't understand it at all. It's like a new thing to us that's saying person A is more risky than person B and our knowledge is no, actually person B is more risky. So we had a whole discussion with the Qlik Sense team about what goes into it and what doesn't. Now for us, when we're managing people, we look at nine pathways of their life that could help them turn away from offending. So you look at accommodation needs, their drugs and alcohol needs, their mental health or physical health, children and families, finances and there's some specific pathways for women and sex workers…we monitor people so closely that the slightest thing like the breakdown of a relationship could cause them to reoffend. Now that breakdown in the relationship isn't going to go into Qlik Sense because it's not a crime, it's an intelligence report and Qlik Sense doesn't pick up intelligence. So we were quite frustrated by that at the beginning. (inspector)

She went on to note: 'once we accepted … that it wasn't the be-all-and-end-all but it was a tool, then it became much more effective for us.' (inspector) In another interview it was pointed out: 'what the model's doing is very quickly saying hey, there may be some risk here but that's for that practitioner or that local team, it's them that make the informed decisions of who they're going to prioritise and what we're going to resource, and not the model.' (manager) In this context, the system was referred to as a 'triangulation tool' used in addition

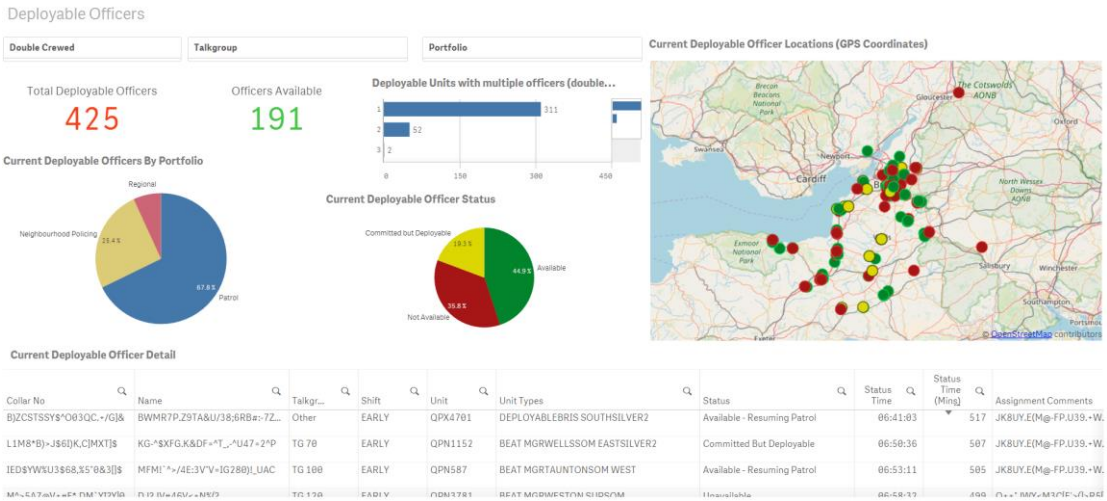
to referrals and intelligence reports, and there 'entirely to support a professional judgement' (manager)

In particular, it was noted that the model is used to provide an indication of individuals that might need management, and to assess what team would be best suited. One inspector explained the traffic light system they use for making this decision:

> we risk assess prolific acquisitive crime offenders using a traffic light system, so we have red, amber and green. The red offenders are managed by a specialist team and the amber offenders are managed by the neighbourhood team because they need less attention. So if we're moving someone from red to amber because they are doing well, that gives us an opportunity to bring on another red one, the team should be looking at QS and saying who's scoring high, do we need to scope them for consideration of them coming onto our scheme? (inspector)

In that sense, the use of the system was explained as being about 'focusing us in the right direction' (inspector), particularly in a context of limited resources. In practice, this means that the system 'highlights it to you that actually you need to get this person into custody sooner rather than later.' (coordinator) In cases of domestic abuse, the system has been used to identify the top 15 offenders who are most at risk of committing a serious abuse offence based on matching characteristics, and engaging other agencies and teams on that basis as part of pre-emptive measures. The system now also tracks the pathways of management for offenders in order to be able to in future evaluate the 'success' of measures taken for lowering risk scores of those people being managed by a particular team.

Moving away from individual risk, Qlik Sense is also used 'to forecast demand, predict demand' and to identify 'command and control incidents that haven't currently been classed as a recordable crime but have all the attributes that they should be.' (manager)



Auditing and Safeguards

It was noted in interviews that the Avon & Somerset Constabulary leadership carried out some consultations with the local community through council meetings in addition to public engagement through media when implementing Qlik Sense. In addition, it was noted that users were surveyed in terms of the proportions that are using it and what they think of it and that when the industry watchdog HMIC visits Qlik Sense 'will typically get mentioned. So they'll go out and check and test the reality of how it's being used on the ground in these different use cases.' (manager)

The system also went through initial testing 'about the individuals that were being identified and having a bit of a check and test on whether these individuals are the right type of individuals. So we went through all of that process and the best top evaluation of the model accuracy, the precision of the models, the recall of the models.' (manager) The Qlik software is updated and changed every ten weeks.

Finally, usage is all audited, 'what people are using, where, when, what and all that kind of stuff. It's also designed in such a way where people are only given the information they need to use and know and that kind of thing.' (former employee)

Challenges

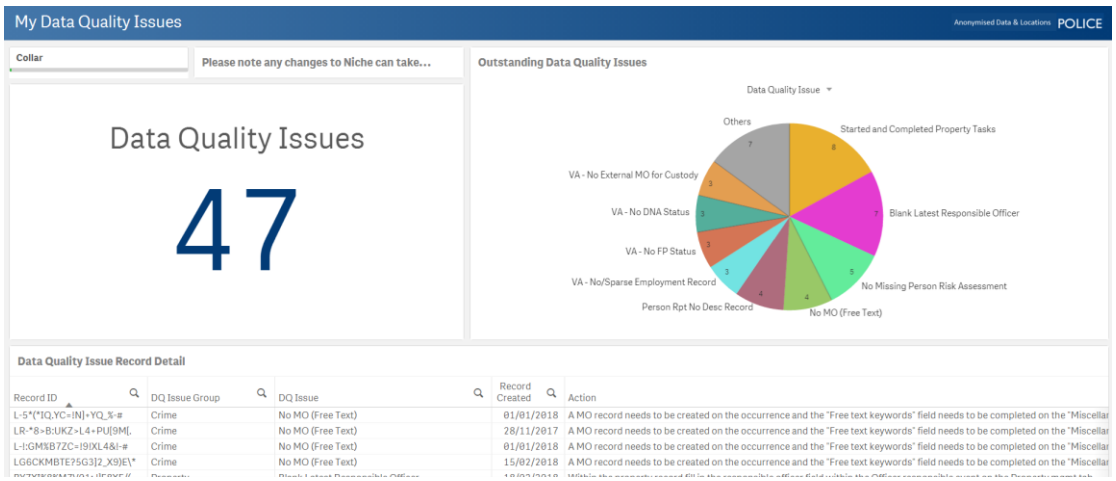
It was noted in interviews that because of previous failures with the implementation of new technologies, a key cultural challenge is that 'a lot of confidence around technology is low.' (former employee) From management's perspective, this relates to a challenge with 'data literacy' that means it is 'working hard on data literacy in terms of people's confidence and ability to engage, interpret and argue data and placing data at the centre point of how people make decisions.' (manager) This was also emphasized in another interview in relation to an unwillingness to share information: 'I think what we've got to do is really, at the same time as pushing out that analytics kind of culture, is really also promote the data literacy culture at the same time as well…because often you find the biggest issues and cases you get is where people don't use information or don't share information and that's often where we get the problems.' (former employee) This also means being more 'proportionate' about sharing information: 'I think if we're expected to continue to reduce our budgets, continue to battle against austerity, continue to make people safer in a way that technology can help us do that, we've just got to be a little bit more proportionate to that…How much do you hear really of information going astray to do terrible things when it's with the police? It's really limited, isn't it?' (former employee)

However, at the same time, amongst frontline staff, a key challenge that was noted in interviews is that too much importance is attributed to the system over and above professional judgement: 'Challenges-wise, I think it's that there are still some people in the organization who believe it is the be-all-and-end-all and professional judgement isn't quite as important. So there will still be people who say…this is what we must do.' (inspector) This was, in part, explained by the question of being able to defend decisions in relation to the model:

> If somebody is shown on our system as being a really high risk of committing, for example, a domestic abuse offence and we say actually that's person A, person B over here we believe is higher risk, even though their [person A] score is higher. So we do something with person B and we don't do something with person A and then person A then goes on to kill someone or seriously injure them, where is the defenceability around that? So I can understand people's thinking in that sense, I really can. (inspector)

In this context, the onus then falls on the frontline staff to record or explain any decision they make that might be at odds with what the system is telling them. In a different interview, the related issue of 'dependency' was highlighted as a potential concern with the system: 'You become reliant on any system and when it goes down, it crashes, then we can still work, it's not going to stop us working but you can become too dependent on technology sometimes.' (coordinator)

This becomes pertinent in relation to the model's accuracy and the challenge of data quality: 'the model doesn't get it right every time' and when it comes to data quality issues, 'we use record management systems that aren't always completed or as accurate as we would like' (manager). In another interview it was similarly noted that being able to interrogate risk scores is important in light of accuracy questions: 'if someone has got a particularly high score, we will look at what's given them the high score and drill in to make sure the data's correct but it isn't always. For example, it might be a data quality issue where someone is identified as high risk because they were previously linked to a murder or attempted murder and actually they were eliminated from that murder.' (coordinator) In response to this, Avon & Somerset police have developed an app to try and capture data quality gaps, 'right down to an individual officer level. So we can track data quality issues over time.' (manager) In one interview it was outlined how this has led to an emphasis on 'personal responsibility' with regards to 'the quality of the information going in': 'we've got thousands of staff inputting records and training and all of these things can lead to data errors. So what we've done is we've utilised Qlik as a system to pull out those errors and actually put them in the officers' domain.' (chief inspector)



A final challenge highlighted in interviews is the ability to actually act on the risks visualised in the system. Whilst the system identifies different aspects, it is another question what plans are in place to manage those different aspects, something that was noted as still being a process (‘we’re getting there’) that relate on ‘the changeover from not having the system before and people getting used to putting things in place.’ (coordinator)

Software case study: Experian Mosaic

Experian is one of the "Big Three" consumer credit-checking agencies, along with TransUnion and Equifax. It employs 17,000 people in 37 countries and, in 2014, was earning almost 5 billion pounds each year (Experian Global, 2014; Experian, 2018b). As well as credit-checking, Experian provides analytics and marketing services to a wide range of businesses, as well the public sector. They have a large, high security data centre outside of Nottingham, UK, and many more data centres around the world (Robson, 2005; Khan, 2017). They are a 'partner company' working with the Government Digital Service (GDS) as part of 'GOV.UK Verify', an identity verification system developed for use by government authorities (GDS, 2018).

There have been a few significant controversies surrounding Experian and its handling of data. In 2014 it was reported that Experian's databases had been involved in 97 'breaches of personal information' (Pagliery, 2014). It was reported in 2013 that Experian had been fooled into selling its data to an identity theft service that sold Social Security numbers, driver's license numbers, bank account and credit card data (Krebs, 2013). In 2015 the personal information of 15 million people was exposed after the company was hacked (Thielman, 2015).

Experian is also one of the 30 'organisations of interest' in the Information Commissioner's Office's (ICO) investigation into Cambridge Analytica and Facebook, relating to the use of data analytics in political campaigns (ICO, 2018b: 32-34; 42; Hern, 2018).

A significant tool for data analytics in public services is Experian's Mosaic tool. In this section, we will outline this tool introduce geodemographics, its history, its evolving role within the UK public sector, and the influence of large, data-rich businesses. We will highlight questions relating to public accountability for proprietary software, commercial entities marketing their products into the public sector, subjective understandings reified in code, and the various sources data is gathered from for systems like Mosaic**.**

Mosaic

Experian's Mosaic is one of the most prominent geodemographic segmentation systems in use today. Geodemographic segmentation refers to the practise of sorting individuals or households into distinct, multivariate, demographic categories, with primacy given to geography under the rationale that "birds of a feather flock together". Veteran of UK geodemographics and the developer behind Acorn and Mosaic, Richard Webber, and sociologist Roger Burrows[108], in their 2018 book, *The Predictive Postcode: The Geodemographic Classification of British Society* - which argues for the further use of geodemographic methods in social research - give the following definition:

> Put simply, geodemographics is a branch of social analysis that recognizes that where you live matters to any understanding of your values, behaviour and choices as a consumer. It uses data from various sources to place each citizen

[108] Of 'The Coming Crisis of Empirical Sociology' fame (Savage & Burrows, 2007).

> into a category according to the type of neighbourhood in which he or she lives. (Webber & Burrows, 2018: xxii)

At one time or another, Mosaic classifications have existed for at least the following territories (Experian 2007; 2010):

- Australia
- Austria
- Belgium
- Canada
- China (Beijing, Guangzhou, Shanghai)
- Czech Republic
- Denmark
- Finland
- France
- Germany
- Greece
- Hong Kong
- Israel
- Italy
- Japan
- New Zealand
- Norway
- Republic of Ireland
- Romania
- Singapore
- Spain
- Sweden
- Switzerland
- The Netherlands
- UK
- USA

Mosaic divides populations into distinct "Types", which are combined into thematic "Groups". The latest UK version of Mosaic contains 15 Groups and 66 Types for segmenting the UK population. Fig. 1 shows a selection of these from Mosaic Public Sector. Fig. 2 shows an example of some of these categorisations mapped onto a suburban area, taken from a business-targeted Mosaic brochure. Experian develops its taxonomy of Groups and Types through the analysis of large amounts of data, in part through the use of *cluster analysis* (Webber & Burrows, 2018: 42-43; 78-79). We discuss these sources of data further below.

These categorisations - optimised at the level of (in pre-2014 versions of Mosaic) households and (post-2014) individuals - are the backbone of the Mosaic system (Webber & Burrows,

2018: 85). They purportedly allow for more direct marketing to specific demographics of people than earlier techniques, with the shift to optimising at the individual level being a capacity enabled by the rise of 'big data' (Ibid.). The phrase "people-based marketing" is one often thrown around in these contexts (Grieves, 2017).

These demographic segmentations are used by Experian's customers to guide marketing, distribution, development, and all manner of other business decisions. For example, a company selling private, premium end-of-life care within a certain geographic area might seek to use Mosaic to identify the addresses of all households or individuals under the Type *Diamond Days* - described by Mosaic as "Retired residents in sizeable homes whose finances are secured by significant assets and generous pensions" - and market directly to them. Or perhaps a budget food retailer seeking to open a store in a new city could use Mosaic to identify areas in the city with the highest concentration of Types under the Group *Municipal Challenge* and use that information to choose an optimum location for their store, perhaps using a map similar to that in Fig. 2.

Fig 1. *A selection of the latest iteration of Mosaic's 15 Groups and 66 Types, taken from a brochure on Mosaic Public Sector (Experian, 2016: 6-8)*

M Family Basics	M53	Budget Generations	Families supporting both adult and younger children where expenditure can often exceed income
	M54	Childcare Squeeze	Younger families with children who own a budget home and are striving to cover all expenses
	M55	Families with Needs	Families with many children living in areas of high deprivation and who need support
	M56	Solid Economy	Stable families with children renting better quality homes from social landlords
N Vintage Value	N57	Seasoned Survivors	Deep-rooted single elderly owners of low value properties whose modest home equity provides some security
	N58	Aided Elderly	Supported elders in specialised accommodation including retirement homes and complexes of small homes
	N59	Pocket Pensions	Elderly singles of limited means renting in developments of compact social homes
	N60	Dependent Greys	Ageing social renters with high levels of need in centrally located developments of small units
	N61	Estate Veterans	Long-standing elderly renters of social homes who have seen neighbours change to a mix of owners and renters
O Municipal Challenge	O62	Low Income Workers	Older social renters settled in low value homes in communities where employment is harder to find
	O63	Streetwise Singles	Hard-pressed singles in low cost social flats searching for opportunities
	O64	High Rise Residents	Renters of social flats in high rise blocks where levels of need are significant
	O65	Crowded Kaleidoscope	Multi-cultural households with children renting social flats in over-crowded conditions
	O66	Inner City Stalwarts	Long-term renters of inner city social flats who have witnessed many changes

Fig. 2 *An example of Mosaic categories overlayed onto a map of a suburban area. (Experian, 2018a)*

Suburban diversity:
While suburban areas can appear on the surface to be a bland monotone area of identikit housing, Mosaic can peer inside. It will reveal all the different types of household that reside there, with their various life-stages, marital status, household compositions and financial positions, as well as taking into account the variations between estates as a whole (cooler colours – better off vs. hotter colours – less well off).

Mosaic Public Sector is marketed as providing 'a pin-sharp picture of the people you need to reach, reflecting the latest socioeconomic trends', using 'the latest analytics techniques' to condense 'over 850 million pieces of information across 450 different data points … to identify 15 summary groups and 66 detailed types that are easy to interpret and understand' (Experian, 2016: 2). Experian call Mosaic Public Sector their 'most comprehensive cross-channel segmentation, built for today's hyper-connected world' (Ibid.).

Whilst Mosaic is widely used within the UK public sector, it is primarily a tool of 'commercial sociology' (Burrows & Gane, 2006). Beyond differences in marketing and customer support, it is unclear if there are any meaningful differences today between the public sector and commercial facing versions of Mosaic in terms of the raw functioning of the classifications and their accompanying visualisations and dashboards. Experian's commercial facing marketing materials for Mosaic state, 'Behind every customer is an individual. Mosaic means you can start treating them that way' (Experian, 2018a: 2). In their leaflet for Mosaic Public Sector they claim, 'Each citizen is an individual. Mosaic Public Sector enables you to treat them in that way' (Experian, 2016: 2).

In the 2003 version of Mosaic, demographic labels differed between the business and public sector marketed versions, to be 'more acceptable to those working in the public sector' (Webber & Burrows, 2018: 113).[109] For example, *Metro Multiculture* became *High density*

[109] Webber & Burrows comment multiple times on how they perceive 'political correctness' to have influenced perceptions of and changes to Mosaic classifications (2018: xxvi; 113; 132). They are not explicit on if they see this as a positive or a negative thing, but their comments elsewhere in the book on what they see as sociology's "overall aim ... to quantify the separate effects of 'class', 'gender', 'race' or whatever variations in some dependent variables of interest ... [for example,] 'voting', 'health' or 'attitudes'" (Webber & Burrows, 2018: xxvii), which they are very critical of, appear to suggest that they take some issue with the explicit politicisation of language. They partially engage with criticisms of the labelling of Mosaic categories in a section titled 'Cluster

social housing, mostly in inner London, with high levels of diversity, and *Welfare Borderline* became *People living in social housing with uncertain employment in deprived areas* (Ibid.). As of the 2014 version of Mosaic, 'although different versions are marketed to the two user groups, the Group and Type labels are now the same for both' (Ibid.: 121). Groups are ordered differently in the latest brochures for the commercial and public sector versions of Mosaic, but beyond that there appears to be little divergence (Experian, 2016; 2018). A housing activist we interviewed for this project, speaking of the uses of Mosaic by the private developer Lendlease who were working with their local authority, said

> It's really stigmatising and it ignores so much else that's going on, so much internal difference, so much shading off. ... That's a big part of how private developers see the world, so the idea that the owners are in one place and the tenants are in another place, that the owners have a collective personality and the tenants have a collective personality and that's what we're seeing here with Experian. ... [The developers are presented with] 66 different [Mosaic] household groups ... you're presented with a picture of usually a man and a woman and "this is what these people look like, this is where they live, here's what their tastes are". So the whole population of an area is reduced down to one couple, which is really bad.

The apparent lack of difference in baseline functionality between the commercial and public sector versions of Mosaic highlights the issue of privately developed software, produced for commercial means, being placed into public sector contexts. This is a recurring theme throughout this project - either with companies remarketing their products produced for the commercial market for the public sector, or public sector organisations purchasing and deploying solutions developed for the market.

Having covered the basics of Mosaic, we will now explore a little of the history and methodological concerns behind Mosaic and geodemographics more generally.

Geodemographics

Geodemographics can be traced back to Victorian social reformer Charles Booth's 1899 'Poverty Map of London'[110] which used reports of school board visitors as its primary source of data, combined with census data, and Booth personally inspecting neighbourhoods (Webber & Burrows, 2018: 32). The pioneering of computer-based geodemographic systems is generally attributed to Jonathan Robbin who, after serving as a faculty member of the Department of Sociology at New York University in the 1960s, left the academy to apply his methods in other areas. His work was to become the first modern geodemographic system, PRIZM (Potential Rating Index for ZIP Markets). Initially the development of this system was funded by the US Department of Housing and Urban Development to target housing grants

labels offend the sensibilities of some social scientists' (Ibid.: 107-113), but they do not take these engagements very far, nor do they engage with arguments relating to social justice and language.

[110] http://mappinglondon.co.uk/2011/the-booth-poverty-map/; https://booth.lse.ac.uk/map/14/-0.0692/51.4859/100/0

to cities with a history of rioting but was later bought up for commercial use (Burrows & Gane, 2006: 43-44).

This trajectory from the public sector to private enterprise - and, as this project is concerned, back to the public sector - is echoed in the UK. Around the same time, working for the Centre for Environmental Studies - a partially government funded thinktank which closed in the 80s - Richard Webber developed the 'Classification of Residential Neighbourhoods' (CRN), the first geodemographic classification of the entire UK, which had a resolution at the level of electoral wards (Webber & Burrows, 2018: 48; 73). In 1979 Webber was recruited by CACI, a US consulting organisation, to enhance the resolution of CRN down to the level of the 'enumeration districts' used by the census. This system was rebranded as 'Acorn' (A Classification of Residential Neighbourhoods) and a descendent of this system is still in active use today, including by local authorities we encountered during our research. Later, Webber went on to found Experian's[111] micromarketing division, out of which came Mosaic (McElhatton, 2004; Webber & Burrows, 2018: 78). The Royal Mail had launched the UK postcode system in 1974 and Mosaic took advantage of this, offering a finer resolution than the contemporary iterations of CRN and Acorn. Mosaic was also the first UK geodemographic system[112] to incorporate non-census data sources. These sources initially included transactional data sources such as county court judgements, property values, property sales information, 'electors with names from different ethno-cultural backgrounds'[113] (Webber & Burrows, 2018: 78).

In their book, Webber and Burrows use Mosaic as their go to example of geodemographic software. They define geodemographics as,

> a branch of social analysis that recognizes that where you live matters to any understanding of your values, behaviour and choices as a consumer. It uses data from various sources to place each citizen into a category according to the type of neighbourhood in which he or she lives. (Webber & Burrows, 2018: xxii)

They are keen to distinguish geodemographics from other forms of social analysis that may categorise people 'on the basis of their own personal characteristics, such as age, gender, ethnicity and so on' (Ibid.: 8);

> In terms of statistical methods, the feature that distinguishes a geodemographic profile is that it uses what are referred to as *multivariate*

[111] Experian was then named CCN.

[112] MAP, a Dutch system, appears to have been one of the first neighbourhood taxonomies to incorporate transactional data. MAP was produced by Wehkamp, a subsidiary of Great University Stores, of which Experian (then named CCN) was also a subsidiary (Webber & Burrows, 2018: 78)

[113] Webber is known for his more recent work on using surnames to profile and target consumers, including attempting to determine their ethnicity (Webber, 2007). Perhaps this is the same software briefly mentioned by the Information Commissioner's Office in their report in the wake of the Cambridge Analytica scandal, claiming a number of UK political parties used software to attempt to determine voters ethnicities (ICO, 2018a: 31).

> categories. Demographic categories are multivariate in that the set of variables used to construct them typically represents different dimensions of social character. (Ibid.)

Through the lense of this method, geography and locality are given primacy. To paraphrase the epigraph Webber & Burrows choose to open the first chapter of their book, 'birds of a feather flock together'.[114] Whilst remaining critical of sociology's focus on 'personal characteristics' they do stop short of calling for the abandonment of their use (Webber & Burrows, 2018: xxiii).

Webber & Burrows make the argument that geodemographics - and, by implicit extension, Mosaic, their go to example - can be fruitfully deployed in the public sector and for non-commercial social research:

> Marketers have long known that such classifications are extremely useful for understanding variations in purchasing habits, tastes, values and so on. We argue that policy-makers, academic social scientists and market researchers are among a number of groups who could similarly benefit from a more extensive engagement with geodemographic modes of analysis. (Webber & Burrows, 2018: xxii)

A controversial supporting point they come back to a few times in their presentation of geodemographics and critique of contemporary social sciences is their view that, referring to approaches which focus on 'personal characteristics',

> such practices lie at the very heart of the failure of the social sciences to provide credible interpretations of the tectonic socio-cultural divisions that now mark post-Brexit Britain and Trump's America. (Webber & Burrows, 2018: xxvii)

Their central claim here is that 'commercial' social science is in some way ahead of more mainstreamed forms of social research and that separation between the two should be broken down. This, in their minds, means the wider adoption of geodemographic methods by the public sector and academic researchers.

In the public sector, our research suggests this is the trend at the local level, although we cannot speak to how far this will go nor are we able to produce a fully comprehensive view of the current use of geodemographic methods in the UK public sector. Webber and Burrows, it is worth noting, highlight a discrepancy between local and central government. They claim central government still appears resistant to the use of geodemographics, whilst local government has appeared more receptive (Webber & Burrows, 2018: 59; 273).

[114] Burrows & Gane, back in 2006, note the repetition 'ad nauseam' of this and another cliché of '"You are where you live" ... throughout the literature as the conceptual justification for geodemographics' (2006: 795).

Public sector use

When deployed in the public sector, Mosaic is often used in conjunction with data already available to public authorities. For example, Kent County Council's Kent Integrated Dataset (KID) uses Mosaic (as well as CACI's Acorn) alongside data accessible by the local authority, to attempt to identify trends and provide insights relating to service usage and provision needs. For example, the KID has been used to attempt to predict how likely individuals are to be unnecessarily admitted to hospital within a year. Its linked data has also been used to attempt to identify GP practises in east Kent that may be over-spending, by combining data on the care costs of individuals using those practises (Kent County Council, 2018a; Abi-Aad, 2016). In response to a Freedom of Information request, Kent County Council described Mosaic as a "socio-political profiling tool" (Kent County Council, 2018b). It is still unclear the exact manner in which Kent CC are using Mosaic and similar tools, but it appears their analysis of their own datasets is augmented or based upon Mosaic's demographic categories.[115] A presentation given by Experian employees in 2014 refers to Mosaic's ability to allow bespoke classifications, built on top of Mosaic's basic Types and Groups (Cresswell et al., 2014: 27).

Webber & Burrows note Experian's efforts to enter the public sector at the turn of the century and the corresponding spread of geodemographics within the public sector in subsequent years:

> [T]he public sector was to become the focus of a concerted Experian push during the early 2000s - local authorities, police forces, hospital trusts and fire and emergency services being their primary targets. By 2006 as many as 100 public sector organizations were attending annual conferences focused exclusively on public sector applications of geodemographic classification. (Webber & Burrows, 2018: 80-82)

They note the variety of government authorities attending these early conferences and some of the use cases discussed:

> The programmes included speakers such as the Director of Transformational Government at the Cabinet Office and the Head of Customer Insight at Her Majesty's Revenue and Customs (HRMC). Attendees could choose between parallel sessions devoted to the applications of geodemographics to health, local government, community safety and education. Other speakers focused on its application to communications campaigns: Nottingham Primary Care Trust, for example, on smoking cessation; Sports England on increasing participation in active sports; and Thames Valley Police on road safety. A key focus of many of the presentations was how to adapt service delivery to reflect the different communication channels that people living in particular geodemographic categories preferred to use. (Ibid.: 82)

[115] Abi-Aad, 2016 gives a short introduction to the KID. This system is also a case study within this report.

Another contemporary example of the use of Mosaic by UK comes from the London Fire Brigade (LFB). Their 'incident risk analysis toolkit' (iRAT) is a collection of statistical analysis techniques that the LFB uses to attempt to identify households in London at risk of fire and which may benefit from intervention. LFB took the data at their disposal on historical fires and overlaid it with Mosaic's groups within London's boroughs. By cross-referencing demographic categorisations with data on previous fires, LFB attempted to identify the sorts of people who are more likely to suffer from a fire, with a resolution down to the size of electoral wards. These insights were then used to produce a map of areas seemingly with a higher risk of fire, even for where a fire had not previously occurred, based off of correlation with Mosaic's classifications. The top three Mosaic demographics highlighted appear to have been those, under the 2006 version of Mosaic's groups, *Twilight Subsistence*, *Welfare Borderline*, and *Grey Perspectives* (Local Government Improvement and Development, 2010; Experian, 2006). Around 630,000 homes were determined to be within 'priority postcodes', down from the 3.5 million homes that make up London (Local Government Improvement and Development, 2010: 6).

In 2016, Lancashire County Council used the 2016 version of Mosaic to produce an analysis of the demographics within its county (Lancashire County Council, 2016). They published corresponding visualisations, using Power BI, a business intelligence tool by Microsoft (Lancashire County Council, 2017). This provides some good examples of the sorts of more raw insights a local authority may attempt to glean from Mosaic. For example,

> Of the 15 mosaic summary groups, 'transient renters', 'senior security' and 'aspiring homemakers' are the most over-represented in the Lancashire-14 area. The areas with transient renters are mainly small, but densely populated, postcodes within central Blackpool, Morecambe and Preston. (Lancashire County Council 2016)

Data

The question of what data - and from where - Mosaic uses is an important one. Due to the proprietary nature of Mosaic, however, it is a difficult one to answer completely. Experian claim Mosaic uses 'over 850 million pieces of information across 450 different data points' to produce their demographic categorisations (Experian, 2016: 2). As noted above, census data still plays a prominent role. A presentation by Experian employees in 2014 claimed 28% of the data used at the time to build Mosaic UK was based on census data (Cresswell et al., 2014: 10). However, one of the selling points of a system like Mosaic is its use of more contemporary data than that of census, given that ten or more years can pass between UK censuses. Fig. 3 and 4 show the further sources of information listed in the same 2014 presentation.

It appears Experian's ConsumerView is utilised by Mosaic, being paired up with higher-level, geography related data. Experian claim,

> ConsumerView provides a single, definitive and consistent view of the UK adult population including contact information across postal, email, mobile and display channels, linking to a broad and accurate range of demographic, socio-

> economic and behavioural characteristics on each adult and household in the UK. (Experian, 2017: 1)

Experian claim to have information on 49 million adults (Ibid.). According to the census, the UK population was 63,182,000 in 2011 (ONS, 2012: 2). Experian claim to have data on 25 million households, 30 million email addresses, 19 million mobile phone numbers, and 10 million landline numbers (Experian, 2017: 2). Furthermore, they write, after quoting these numbers: 'Online display advertising and our match rates across Facebook (50%) and Twitter (35%) mean you can have a richer view of individuals across every channel' (Ibid.).

It appears Experian use a mixture of open data, data-scraping, and agreements with third parties to gain access to new data. However, a comprehensive, public record of the data they use is not available, given Experian's commercial status that means this information is proprietary.

A noteworthy source of data briefly mentioned is Emma's Diary - who trade as Lifecycle Marketing (Mother and Baby) Ltd. - used for data relating to children (Cresswell et al., 2014: 8). Emma's Diary is a website which provides advice and information to parents.[116] As part of the Information Commissioner's Office's 2018 investigation into the use of data analytics in political campaigns - in the wake of the controversy surrounding Facebook and Cambridge Analytica (The Guardian, 2018-) - the ICO has issued Emma's Diary with a £140,000 fine under the Data Protection Act 1998 because of a 'serious contravention of the first data protection principle' relating to the fair and lawful processing of data. The ICO claim that Emma's Diary did not sufficiently notify their users of how their data would be used, with this particular case coming to light because the UK Labour Party, during the investigation, told the ICO that they had been sold data from Emma's Diary. The Labour Party gained access to this data via Experian, so it would appear to be the same data hitherto mentioned in the Mosaic presentation (Kelion, 2018; ICO, 2018a: 24; 2018c: 9).

We only know a limited amount about the data used by Mosaic, nevermind the intricacies of the analytical methods applied to that data, and the further discretion applied during the "art" (Webber & Burrows, 2018: 85; 131) of geodemographics, used to arrive at Mosaic's classifications. Knowledge of these sources and methods is the intellectual property of Experian. A housing activist we interviewed for this project claimed this was a barrier to their work, with residents unable to interrogate the data behind a planned development because they were the property of the developer, Lendlease, and Experian, via Mosaic.

[116] https://www.emmasdiary.co.uk/

Fig. 3 *A selection of data sources used for the 2014 version of Mosaic (Cresswell et al., 2014: 7).*

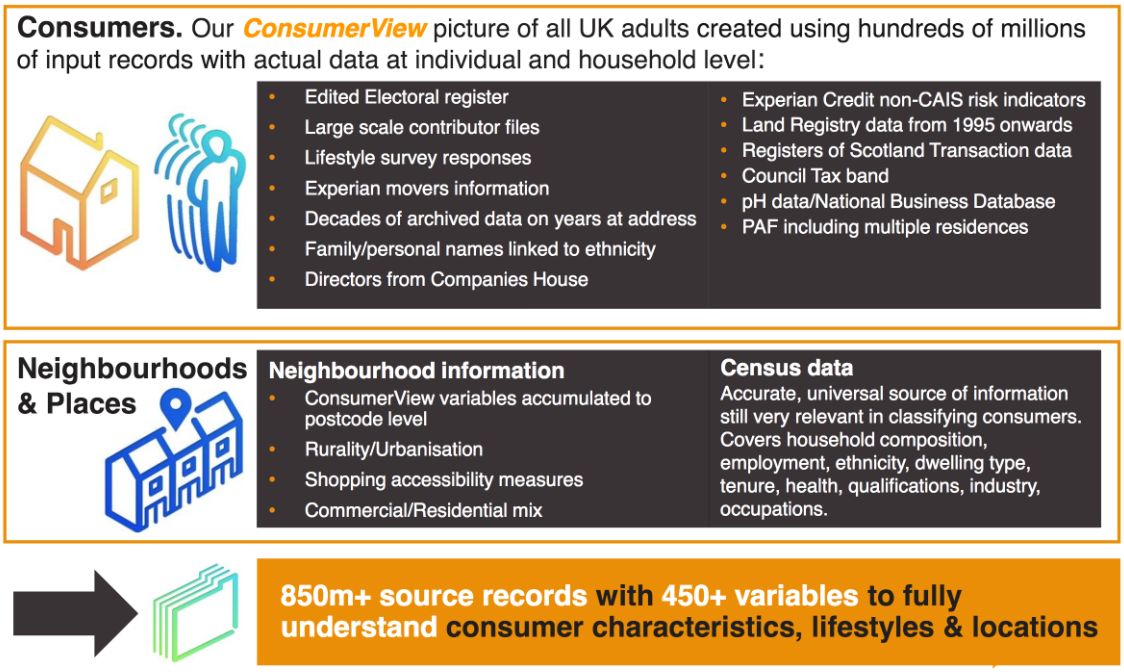
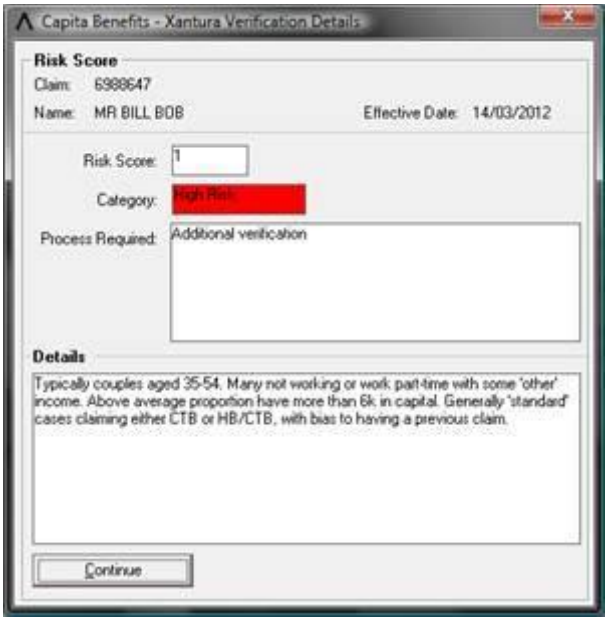


Fig. 4 *Further data sources used for the 2014 version of Mosaic (Cresswell et al., 2014: 8).*

- **ConsumerView Variable Improvements** using new data:
  - **Tenure & Property** - Rightmove and National Register Social Housing
  - **Children** using Emma's diary
  - **Northern Ireland detailed property data** available for first time
  - Improvements from **refreshed census data** used in calibrations.
- **Higher Education Statistics Authority (HESA)** college and university students database linked to home & term time postcodes
- **Transience & attractiveness** - Experian movers data indicates how populations are changing
- **OS Streetview** - Mapping data about property area, ratio of gardens to buildings, cul-de-sacs, sea/part/lake view
- **Census 2011** - a new snapshot of entire UK population
- **Accessibility** to shopping centres, high streets; distance to motorways, schools, railway stations, coast, GPs, bus stops….
- **Level of urbanisation/rurality** based on population densities
- **Open Data** - Police.UK crime data, GCSE results, Gas & Electricity consumption, DWP child benefits, tax credits, income support by Census LSOA areas
- **Broadband speeds** (OFCOM) - Postcode level

Company profiles

Xantura



A screenshot of Xantura's Risk Based Verification (RBV) system[117]

Founded in 2008, Xantura provides data sharing and analytics services to over 70 public sector clients across the UK. They list their key areas of focus as, "improving outcomes for vulnerable groups, protecting the public purse and, helping [their] clients build 'smarter' business processes".[118] They provide systems relating to the Troubled Families programme, fraud and error detection, and children's safeguarding.

Xantura have trialled their Early Help Profiling System (EHPS) in at least the London Boroughs of Hackney, Newham, and Tower Hamlets, as well as Thurrock Council. The system 'translates data on families into risk profiles, sending monthly written reports to council workers with the 20 families in most urgent need of support. ... The risk score created for each family comes in 40 bands, and is put through a "natural language generator" to give summaries outlining why each family passed the risk threshold.' The development of this system was funded by London Ventures, a program run by Ernst & Young - a multinational professional services and accounting firm - and a consortium comprised of London's 32 borough councils and the City of London. It is claimed that the system can save '$160,000 [~£124,000] by replacing human-conducted screenings with an automated system' (Graham, 2017). See our London Borough of Hackney case study for more information on this system.

Xantura provide a Risk Based Verification (RBV) system for the automated detection of "fraud and error", which they claim is used by more than 40 local authorities in the processing of benefits claims. The system integrates with 'either the Capita or Northgate Housing Benefit

[117] https://www.xantura.com/focus-areas/risk-based-verification [accessed 15th November 2018]

[118] https://www.xantura.com/ [accessed 22nd August 2018]

applications and or the following e-Claim solutions: Northgate, Capita, Team Netsol (BECS), Victoria Forms, Web Labs.'[119] A Medway Council contract suggests this system costs around £45,000.[120] A Department for Work and Pensions (DWP) document circulated to Local Authorities 9th November 2011, highlighting the claimed advantages of the use of RBV and pointing Local Authorities towards its use on a voluntary basis, defines it as follows:

> RBV is a method of applying different levels of checks to benefit claims according to the risk associated with those claims ... RBV assigns a risk rating to each HB/CTB [Housing Benefit / Council Tax Benefit] claim. This determines the level of verification required. Greater activity is therefore targeted toward checking those cases deemed to be at a highest risk of involving fraud and/or error. (Department for Work and Pensions, 2011: 3-4)

A 2017 Chichester District Council document details plans to implement Xantura's RBV system, including how they plan to treat benefit claims differently according to the risk score assigned to them, such as by requiring different documentation (Chichester District Council. 2017). Other companies, including CallCredit and Capita, offer a similar RBV service to Councils.[121] Xantura also provide a complimentary system for changes in circumstances of benefit claimants.[122]

Around 2010-11 Xantura ran a trial in North Lincolnshire for their Victims and Vulnerable Persons Index (VVPI) which "was launched as an early warning system about people at risk of systematic attack or abuse by neighbourhood gangs" and used data from various local government agencies, including Lincolnshire Police, National Probation Service, NHS (local hospitals and GPs), Youth Justice Board, Lincolnshire County Council, Department for Work and Pensions, and Safer Neighbourhoods North Lincolnshire. This system was implemented in the wake of the death of Fiona Pilkington and her daughter, who suffered years of anti-social abuse (Government Technology, 2010: 9; ITV News, ~2010).

---

[119] https://www.xantura.com/focus-areas/risk-based-verification [accessed 22nd August 2018]

[120] https://procontract.due-north.com/ContractsRegister/ViewContractDetails?contractId=8fb845cf-b0ef-e711-80e7-005056b64545 [accessed 22nd August 2018]

[121] https://www.callcredit.co.uk/markets-served/publicsector/risk-based-verification; https://csssecure.capita-software.co.uk/cmsstorage/capita/files/52/5298305a-9de0-4623-9f25-3564edbb6422.pdf [both accessed 22nd August 2018]

[122] https://www.xantura.com/focus-areas/risk-based-verification [accessed 22nd August 2018]

Callcredit

Graphics from the "Public Sector Solutions" section of CallCredit's website[123]

Callcredit advertise themselves as providing software and services in the areas of "Credit Risk & Affordability", "Fraud & ID", "Collections & Recoveries", "Customer Experience Data & Decisioning", "Consumer Marketing Data", "Digital Marketing", and "Retail Location Planning", and assist businesses with international expansion.[124] However, they have been most widely known for their consumer credit reporting service, Noddle.[125] In 2018 the company was acquired for around £1 billion by TransUnion, one of the "Big Three" consumer credit reporting agencies (Transunion, 2018).

Callcredit, similar to Capita and Xantura, offer a Risk Based Verification (RBV) system service to councils processing Housing and Council Tax benefits claims.[126] (See our section on Xantura for a description of RBV) They describe their system as

> Fully compliant with DWP's HB/CTB S11/2011[127] guidance and available through all major back office and e-Forms software, Callcredit's solution has already been adopted by over 70 Local Authorities throughout the UK. The solution provides an individual risk rating for all claims, based on the likelihood of it being fraudulent or erroneous. Benefit Officers are then empowered to apply a manual verification process which is appropriate and proportionate to the risk posed by each claim.[128]

Their system uses data from credit reference agencies:

---

[123] https://www.callcredit.co.uk/markets-served/publicsector [accessed 15th November 2018]

[124] https://www.callcredit.co.uk/ [accessed 22nd August 2018]

[125] https://www.noddle.co.uk/ [accessed 22nd August 2018]

[126] Callcredit suggest that RBVs will be phased out with the rollout of Universal Credit: "With strong indications of a delay in the Universal Credit programme affecting the potential inclusion of housing costs, Local Authorities around the UK are still selecting Intercept RBV for Housing and Council Tax Benefits." - https://www.callcredit.co.uk/markets-served/publicsector/risk-based-verification [accessed 22nd August 2018]

[127] Department for Work and Pensions, 2011

[128] https://www.callcredit.co.uk/markets-served/publicsector/risk-based-verification [accessed 22nd August 2018]

> To ensure maximum accuracy for Benefit assessors, Callcredit is now incorporating business intelligence from credit reference agencies to inform the decision making process even further. Using this enhanced data the higher risk claims are automatically exposed to greater scrutiny, increasing the amount of fraud and error detected for those claims.[129]

In 2014 Callcredit acquired Coactiva and incorporated the company into their Public Sector Team, allowing them to supply "bespoke and targeted solutions for public sector organisations, bringing innovative 'Big Data' analytics and business intelligence solutions into everyday use"[130] Their other local government services relate to other areas of fraud detection, debt collection, and profiling.[131] They have developed systems to assist local authorities with rolling out Universal Credit, utilising their experience in fraud detection and demographic profiling.[132] Their 'ThreeSixty Online' system claims to be able to "verify citizen identities, trace debtors, assess personal financial circumstances and screen employees".[133]

Callcredit's demographic profiling tool is named ThreeSixty CAMEO and claims to be able to assist with profiling and identifying 'the individuals most affected by Welfare Reform". They claim it can:

> - Understand ability and attitudes for aspects such as technology usage or financial literacy
>
> - Establish household composition, ethnicity, educational attainment level and age profile
>
> ...
>
> - Personalise your campaigns and strategies to specific needs and concerns of citizens[134]

For a more detailed look at a comparable demographic profiling system, see our Experian Mosaic case study.

As with Equifax and Experian - the other two of the "Big Three" consumer credit reporting agencies - TransUnion, of which Callcredit is now a part, have a history of complaints from customers using their credit reporting services relating to inaccuracies within their data and delays in correcting mistakes (Hussain, 2018).

---

[129] Ibid.

[130] https://www.callcredit.co.uk/markets-served/publicsector & https://www.callcredit.co.uk/press-office/news/2015/04/coactiva-appointments [both accessed 22nd August 2018]

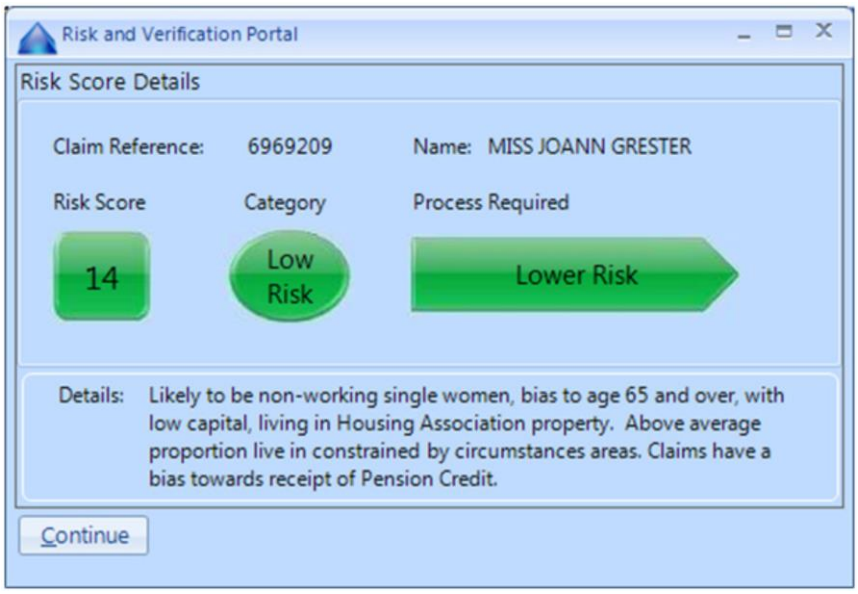
[131] https://www.callcredit.co.uk/markets-served/publicsector/local-government [accessed 22nd August 2018]

[132] https://www.callcredit.co.uk/markets-served/publicsector/intercept-radar [accessed 22nd August 2018]

[133] https://www.callcredit.co.uk/markets-served/publicsector/threesixty-online [accessed 22nd August 2018]

[134] https://www.callcredit.co.uk/markets-served/publicsector/threesixty-cameo [accessed 22nd August 2018]

Capita



A screenshot of Capita's Risk Based Verification (RBV) system[135]

Capita is a large public outsourcing company. Over the years it has taken on many public contracts including electronic tagging of offenders, collecting the BBC licence fee, implementing London's congestion charge, administrating winter fuel payments for pensioners, providing supply teachers to schools, and software for booking driving tests (Travis, 2014; Taylor, 2006)

Like Callcredit and Xantura, Capita provide a Risk Based Verification (RBV) system to semi-automate the processing of Housing Benefit and Council Tax Benefit claims. See our section on Xantura for more detail on this method. Of their system, Capita write: 'Capita's Risk and Verification Portal uses sophisticated data analytics and predictive modelling techniques to calculate the 'risk' or probability of a claim containing errors.'[136]

A promotional video on Capita's website, under the branding "One", outlines their claims that their products and services can help local authorities link up public and community services to gain a more comprehensive view of service needs, facilitating appropriate interventions.[137] A brochure for Capita One, aimed at the public sector, advertises a very wide array of services, including:

> The complete picture of a child, young person and family ... Comprehensive property and household information with extensive integration and automation ... 360° view of tenants and properties, with mobile estate

135 https://csssecure.capita-software.co.uk/cmsstorage/capita/files/52/5298305a-9de0-4623-9f25-3564edbb6422.pdf [accessed 22nd August 2018]

136 https://csssecure.capita-software.co.uk/cmsstorage/capita/files/52/5298305a-9de0-4623-9f25-3564edbb6422.pdf [accessed 22nd August 2018]

137 https://www.capita-one.co.uk/about-us [accessed 22nd August 2018]

> management and automation ... Support integration with health and help people live independent lives ... Secure, convenient payments for rent, council tax, waste services and other income, as well as expenditure such as grants and benefits ... Fast access to information for staff and a better experience for customers with streamlined electronic document management ... View information from police and 3rd sector services ... Citizens apply online for benefits, school places and early years funding as well as pay rent and log repairs ... Support multi-agency working with shared information amongst appropriate professionals[138]

The same brochure contains endorsements of Capita's services from Calderdale Council and White Horse District Council. Capita's services are used widely across the public sector, from national departments down to local authorities. Some examples include Newcastle City Council's children's social services (Newcastle City Council, 2017: 5-8); Durham County Council maintain a "county-wide Pupil Database" using Capita One (Durham County Council, 2013: 4); Devon County Council appear to be in the process of implementing Capita One Education, which Capita claim on their website "can help you monitor the safeguarding of children across your authority and measure the performance of vulnerable children".[139] The interactive tool we have produced as part of this project can be used to attempt to identify more local authorities using Capita's services.

Among some commentators and MPs, Capita has received significant criticism. Concerning Capita's poorly performing Defence Recruiting System, which led to armed forces applications nearly grinding to a halt, Conservative Party MP Mark Francois used the moniker for Capita popularised by *Private Eye* - "Crapita" - in the House of Commons[140] (Corfield, 2018; Hill, 2018). In 2014, a leaked letter to The Guardian revealed that the government had sent in civil servants to assist Capita with a backlog of face-to-face Personal Independence Payment (PIP) application assessments for terminally ill, sick, and disabled individuals, which Capita had been contracted to carry out but had fallen drastically behind on. In some cases, waiting times for assessments were so long that people with terminal conditions died before receiving any money (Malik, 2014). In 2016 some reports suggested that 61% of claimants who appealed against a PIP decision made by the Department for Work and Pensions (DWP), based on Capita's assessments, won their case at tribunal. Tim Farron, then leader of the Liberal Democrats, claimed Capita "are just driven by a profit motive, and the incentive is to get the assessments done, but not necessarily to get the assessments right" (Walker, 2016).

---

[138] p.1: https://www.capita-one.co.uk/sites/default/files/2018-03/One-Overview-Brochure-WEB.PDF [accessed 23rd August 2018]

[139] https://oneonline.devon.gov.uk/CCSEnterpriseOnline_LIVE/default.aspx & https://www.capita-one.co.uk/product-and-services/one-education-and-childrens-services/education-services [both accessed 23rd August 2018. At the time of writing the Devon County Council page claims "The system is being initialised and will be available shortly."]

[140] Much to the amusement of a group of schoolchildren observing from the public gallery (Echo, 2018).

Section V:
Civil Society Perspectives

Civil Society Perspectives

In this section we outline key themes that emerged in our interviews with different groups from across civil society identified as stakeholders in developments of datafication in public services. This includes groups concerned with digital rights, welfare rights, education, policing, and criminal justice (see methodology section for full details of sample). In our interviews we explored questions relating to the extent of knowledge about uses of data analytics in public services, key benefits and concerns, experiences of data harms and the potentials for addressing any concerns. Whilst there was a recognition of the potential benefits of data analytics in public services, prominent concerns emerged in the interviews with regards to current trends in the implementation of data systems. These regarded the extent of data collection and sharing, the potential for bias and discrimination in decision-making, the possibility of targeting, stigma and stereotyping of particular groups, lack of transparency, public knowledge, consent, and oversight, and the limits of regulation to address the overall political context of uses of data systems in public services.

Extent of data collection and sharing

Several civil society groups expressed concerns with the extent of data collection and sharing now prominent in local and central government. There was a perception that there is a 'data maximisation' trend being advanced in which the onus is to collect as much data as possible and increase data sharing where feasible. This was seen as being significant in part because of the often very sensitive data that local authorities hold that raises questions about privacy:

> It's obviously becoming incredibly easy to accumulate, process, analyse data, look for new insights and there's certainly some exciting stuff about that but the non-consensual data, maximisation data process that we currently are seeing as a trend has to stop. It's the only way to really sensible mitigate the risks. As long as the data exists, it's going to have capital value for someone and it's going to be exploited. So I think data minimisation is key. (Big Brother Watch)

> I think data minimisation is important. So obviously collecting the minimum amount of data that's needed and having this idea of privacy by design. So when technology or things that work off data are developed, having privacy hardwired in right from the beginning as a key consideration is essential. (Liberty)

> I know there are laws, you know, data protection laws. But in terms of government or councils handling sensitive data, that needs to be looked at very carefully. (DPAC)

Concerns with privacy were expressed in several ways and also incorporated the question about the extent to which increasing data collection might deter people from seeking certain services. In particular, there was a sense that this would impact on those who feel especially vulnerable or have feelings of distrust towards the state, such as in the case of the National

Health Service sharing data with the Home Office which might influence activities amongst migrant communities:

> For example, people who are worried about their immigration status feeling like they can't access their GP or go to a hospital appointment because even in those areas, you're seeing data sharing that might leave them vulnerable to whatever it is that they're worried about; even deportation in some circumstances. (Liberty)

In addition, a 'data maximisation' trend as mentioned by interviewees was noted as changing the provision of public services by making working practices of frontline staff more data-centric in both their understanding and approach to service-users. This was noted as significantly shifting the focus of the substance of the work, particularly amongst social workers:

> the data being collected for performance management purposes at an aggregate level and what social workers are being asked to do every day in their individual practice…have become linked now. We've ended up with social workers being data collectors for central government which then sometimes takes them away from day to day duties (…) The systems are set up for social workers to collect data as performance management. We're concerned that that can divert the social worker from being able to understand the case because the sort of data that they're collecting, they might be lost in there, the complexities of the case. (Godfred Boahen, BASW)

Related to this, there was a concern with the extent to which the onus on data collection changes how both problems and solutions are defined in terms of data, suggesting a form of data solutionism:

> if you're thinking about the kind of risks of harm to individuals, institutions or the wider community, then I think they will primarily stem from having a senior level conversation about what the problem is, why data and data use and data analytics is the right solution because many of these problems can be solved in multiple different ways, data isn't the only one. The problem of identifying stabbing hotspots, for example, doesn't have to be a data question or not primarily a data question. You could deal with that in multiple different ways by policing differently and so on. Clearly data might enhance that but also many of these problems aren't going to be solved solely by using data. So a clear problem statement, a clear understanding and statement of what the output is and outcome is that's intended and why it is, and how the data is going to contribute to the solution as opposed to moving straight to data as being the solution. (Involve)

Moreover, there was a concern with the way that extensive data collection fundamentally changes the nature of power, creating an inherent asymmetry between state and citizen through information gathering:

> I think it is true that, through digital systems, it's quite easy to almost reduce people to data and once you have that relation, you really create a new situation of power that is different from the power of the state (…) I think the digitisation creates a new…situation where the Government now are the ones to know everything about you, everything about your neighbours and in that situation, the more data that is collected, the more power people have and it's almost axiomatic, you don't have to explain that information is power. (Open Rights Group)

Bias and discrimination

A prominent concern across civil society groups is the issue of bias and discrimination. The issue of how skewed data-sets might create disparate impacts for certain groups is one that has been well-documented in discussions of data-driven decision-making. In our interviews with civil society actors, this concern was prominent because it emphasises the way that different parts of society will experience the turn to data systems in public services differently and that it will be more relevant for some communities who will also be the ones least positioned to influence developments:

> it's not something that the bulk of the population will ever encounter. It's something you only encounter when you are part of a risk group, a risk population. … So it only ever negatively affects a minority of the population and that particularly concerns me because you've got no leverage really to get it changed. (Netpol)

> one of the perceptions they might have is that technology is beneficial in that it's not biased like humans. It can make completely neutral decisions but, of course, that's not true and the way that we use Big Data to train AI means that it will still perpetuate the same biases and the same pre-existing inequalities that we already have in our society. (Liberty)

Whilst these forms of bias and discrimination may not be intentional and may be comprehensively considered by developers and managers in the deployment of data systems in public services, there was a concern expressed about the ease with which it is possible to also create intentionally discriminatory systems once people's data is collected and used for decision-making:

> you can build discriminatory systems that are not openly discriminatory. Many people are trying to fight that and the assumption at the moment is that all those things happened accidentally and I think it's true that right now in the main, they're happening accidentally but you could build them on purpose as well. Once you understand what the proxies are and that you don't need to actually go around asking people whether they are gypsies or black, you can actually design a system to identify who those people are. (Open Rights Group)

Targeting, stigma and stereotyping

Linked to the question of bias and discrimination, several civil society actors interviewed raised the issue of how data systems can lead to forms of targeting, stigmatisation and stereotyping of particular groups or particular activities. This was particularly raised in relation to how 'risk' as a category is interpreted in data systems and how people might perceive such categorisation:

> Because of this kind of quantification and categorisation approach that data analytics actually demands and the use of ever more sensitive data, there are people who will feel sidelined, maligned, judged, stereotyped. I think that's obviously going to be an issue where, for example, parents who are about to have children are judged as at risk through data analytic schemes and that kind of thing. (Big Brother Watch)

> Obviously people on welfare feel particularly targeted, always have done and that seems to be getting worse. But even slightly differently but in the policing context, the way that it's the people on the margins who live in less good postcodes or don't have such a great credit profile who are judged as higher risk. (Big Brother Watch)

> The concerns are at the individual level that you might be targeted for retribution… it was an integrated dataset from a local authority we drew a case study from that was posited around delivering benefit to the public but actually was about trying to identify where council tax wasn't being paid, and you could begin to see how that might make people feel that data is being used to target them. (Involve)

> Stigmatisation is a very real problem and also, in some cases, it's almost like the state will know more about you. You think that you're normal working class, maybe a poor family and suddenly you are being classed as a risk in some way. It's a fundamental question, what right do you have to label people based on something. (Open Rights Group)

These concerns with targeting and stigmatisation were also noted in relation to the permanence of labelling and what rights a person or family has once they have been categorised:

> One thing, there was a new legal gateway established so that the DWP could share data with local authorities in respect of Troubled Families and what that meant was that where a family had been identified as a troubled family or potential troubled family, a marker was put on their DWP [Department of Work and Pensions] data record that identified them as a troubled family. So if they moved off work benefits, they could then be claimed. That marker stayed on and stays on regardless of whether they have been classed as turned round, whether they've exited the programme or other things have changed. So even if somebody's been in work for months and every other issue in the

> house has been solved, they remain on the DWP system as being targeted as a troubled family. (welfare rights activist)

Lack of transparency and public knowledge

Several of the civil society actors we interviewed pointed to a fundamental lack of transparency and public debate about what data systems are implemented in public services, and noted a lack of clarity regarding what oversight mechanisms are in place. This becomes important for handling any effects of data-driven decision-making, particularly as they 'come to engage human rights' (Liberty). Lack of transparency was noted prominently in relation to the criteria for making interventions off the back of data processing. One interviewee stated:

> What's…concerning is the lack of transparency of how that happens. Where is the criteria? How do you pick out person A and not person B for that sort of intervention? And there's all sorts of those interventions going on where that process is clearly about risk analysis, it's clearly about some sort of data processing but it's really unclear how the end result arises. (Netpol)

Without clarity on this as a preliminary measure, it was noted as a difficulty to consider possibilities of oversight and to address any harms that might emerge from such intervention. This was also linked to the prevalence of public-private partnerships that underpins much of the implementation of data systems in public services:

> the first step is that there has to be more transparency about how data is used and how it's processed and where it ends up and what the purpose of it is for. (Netpol)

> The big thing with outsourcing and partnerships, as they call them, is that you lose public access. (Defend Council Housing)

Moreover, lack of transparency was linked to lack of public knowledge about developments in data and the way data analytics is used in public services. This, in turn, was seen as being significant for the possibilities for greater citizen participation and voice in how these systems are developed and used:

> for all the campaign organisations, for all of the watchdogs, commissioners, oversight bodies, European law, national law, ethics, data ethics bodies, etc., we've still got major, major problems and that's because public understanding just isn't… there's a massive gulf between the reality of what's happening in this field and the public understanding of it. (Big Brother Watch)

Lack of consent and oversight

Linked to a concern with lack of transparency surrounding data processing in public services, was a prominent concern with obtaining consent for how people's data is used. Predominantly, this was noted in relation to the extent of actual informed consent as data becomes repurposed and combined in different ways:

> lots of families may well have signed a consent form about their data being shared and being used in relation to the Troubled Families programme or other services, but I don't think any of them or many of them fully understand what they've consented to and they're certainly not making informed decisions. (welfare rights activist)

> the importance of users of anything that's requiring your data, understanding properly what happens to your data, how it can be shared, if it can be shared and how it's stored, how long it's kept for, because often those things are not clearly set out and so you're consenting to something that you have no hope of properly understanding because it's incredibly extensive and full of jargon. (Liberty)

These concerns about informed consent are linked to a general concern about limited public and citizen engagement with how these processes are implemented, including lack of proper oversight:

> I think we need to strengthen the rights of service users to access their data. We need to strengthen their rights because right now, you can ask for a Freedom of Information request but…it takes a long time and there's so many different parts of the local authority who might hold information on you, that it might actually be difficult for you to get that information. (Godfred Boahen, BASW)

This means that questions as to how people become implicated or impacted by sharing data is often obscured, including any consultations that might engage concerns with targeting, bias or stereotyping:

> I think the issue is that things are being introduced so quickly and without adequate oversight and without adequate testing for things like bias and if you look at things like facial recognition, that's being rolled out without any public engagement, very little Parliamentary oversight and also no law. (Liberty)

Limits of regulation

Whilst several civil society groups expressed the need for stronger regulation around the collection and use of data in general, and in the public sector in particular, there was also a concern with the way regulation might be limited in and of itself:

> I think that institutionalising a culture of data use will require some prior work on building capacity and part of that capacity has to be thinking about building a better understanding of ethics and then thinking about how you operationalise it, and that is primarily I think having strong frameworks of data ethics but then building them into management and governance processes, effectively. And some of the stuff around privacy impact assessments gets some of that way but, in many ways, it appears to be a retrospective and tick box exercise as opposed to being something that is organic within the way an organisation works. (Involve)

> It's an ideological decision to cut the size of the state that's behind all of this. So, I don't think data laws are the way through. It's a change of heart of government or a change of government is what's required. The political will is not there to put right these things. Until that happens, the rest isn't going to make any difference. (DPAC)

Politics, not technology

Linked to concerns with the limits of regulation was a more general point made by several of the civil society actors we interviewed that emphasised the political dimension of data analytics, particular as systems are implemented in public services. This moves concerns away from the specifics of the technologies used, and instead points to broader policy agendas that might be enacted through such technologies:

> I don't think it's about the technology, I think we have to take a step back and look at what the technology is doing and is that an okay thing, with or without the technology? Is that what we want our police forces to be doing? And if so, in what circumstances is that okay? In what circumstances is that not okay? (Netpol)

> I think the problem isn't the method, at the moment, with data collection. The problem is that budget cuts have made, particularly councils and the government, try to skew everything to the cheapest possible outcome. So, where there are data gathering, being whatever they may, that's being, if you like, used to determine a pre-determined outcome. (DPAC)

> if used in a beneficial way, very great because the ability that we now have to collate data and to marshal data, the amount that we can know about the society that we're in has increased. There's no doubt about that and this could be used to tailor services to the needs of people. That's not what happens, unfortunately. That's not the direction of policy. The direction of policy is in the opposite direction. (Defend Council Housing)

> all too often data is being seen as the solution to a problem as opposed to thinking about the problem, thinking about what outcome we want and then thinking about how does data contribute alongside all the other interventions we're going to be making towards the solution of this problem? So what you end up getting is this very tech focused, quite alienating, approach to policy as opposed to the tech, in this case data and data analytics, supporting all of the other work that's going on to build the future that citizens want. (Involve)

In carrying out interviews with civil society actors, we can therefore identify a number of key concerns that are not necessarily considered amongst practitioners who are either developing or using data systems for the purposes of delivering public services. These concerns echo themes that have been highlighted in a range of scholarly research and civil society reports as outlined in previous parts of this report.

Section VI:
Workshops

Workshops

Scoping workshop (20 April 2018, London)

As part of the project 'Data Scores as Governance', the Data Justice Lab held a fact-finding workshop on 20 April 2018 to bring together representatives from government authorities, other public sector organisations, academic institutions, think tanks, and civil society organisations. The goal was to a) explore the state of government uses of data analytics, b) investigate challenges and opportunities, and c) offer a space for dialogue between different stakeholders. The workshop was divided into three discussion sessions:

- Data Analytics in the Public Sector: Experiences and Opportunities
- Exploring Data Uses: Research and Reviews
- Civil Society Perspectives: Challenges and Concerns

Each session was opened with short presentations from specific participants, followed by an open discussion.

The first session – Data Analytics in the Public Sector – centred around contributions of attendees involved with uses of data analytics and related systems in the public sector. Issues discussed included the opportunities of linking datasets across a variety of domains, and the use of commercially developed demographic segmentation tools. Experiences with software such as Mosaic (produced by Experian) differed, from praise of its usefulness for public service allocation to claims of inaccuracy. In many cases the use of data analytics is still in trial stage. Data intensive programmes such as Troubled Families were mentioned. Some concerns for local government attendees related to problems of "changing the culture" towards data sharing and public perceptions around government's use of data.

The second session – Exploring Data Uses – focused on research aimed at producing ethical frameworks and models for using data in government, discussed different types of such frameworks and argued for caution and safeguards in the use of data systems. Participants highlighted the need for a more developed, nuanced, and informed public discourse on data and for better public engagement.

The third session – Civil Society Perspectives – discussed concerns by participants engaged with issues of digital rights, privacy and public involvement. It demonstrated the difference in perception across stakeholders regarding the opportunities and challenges, respectively, of data analytics. The use of privately developed analytical tools that aggregate data from a variety of different sources, including consumer, transactional and private communications data, for public service allocation was criticized in particular. Calls for transparency and public involvement re-appeared, as summarized by one participant: "If AI systems are to shape the world, we must also have the opportunity to shape them." Questions were raised regarding public-private partnerships, opaque algorithms, data used in contexts other than the one it was collected for, openness or lack thereof, and public discourse.

The workshop highlighted fault-lines between different participants in the debate, but also a common concern with the public good. It demonstrated the struggle of weighing and prioritizing different aspects and perhaps different understandings of the public good, including the enhancement of public services and public safety through data analytics, the protection of privacy, and safeguards against data-based discrimination. Below we outline some of these themes and controversies in more detail.

Demographic segmentation software

The use of commercially available software – such as CACI's demographic segmentation software, Acorn, analytic services provided by IBM, and the credit rating agency Experian's software, Mosaic – led to repeated controversies. They are used widely across different data analytics projects in the UK public sector, and were seen as accepted and useful industry standards by most government-related attendees. Yet questions emerged regarding data sources, accuracy and transparency, as well as the implications of pervasive data collection and profiling.

Public-private partnerships

A related theme concerned the use of privately developed, privately owned software by public authorities. While public sector participants defended it as optimizing available resources (it is more cost-effective to purchase or subscribe to such tools, rather than develop an alternative in-house) and using the most advanced technology available on the market, developed with resources and expertise that local government does not have access to, other participants criticised a lack of transparency and public accountability, as well as the inclusion of seemingly unrelated types of personal data in some of these systems.

Public engagement & accountability

While there is strong desire for public involvement and a more knowledgeable public discourse, the practices of public engagement have, so far, strong limitations. Concern was expressed about the lack of effective participation models, the lack of interest by parts of the public to get involved, and a scandal-oriented public discourse.

Privacy vs data

While government attendees rejected the existence of a conflict between data analytics and privacy, due to the safety of systems, anony-/pseudonymisation, and public sector care, civil society attendees maintained that pervasive data collection, analysis and sharing was a fundamental risk for citizens and change the shape of state-citizen relations.

Ethical frameworks

A multitude of models and ethical frameworks relating to data analytics and AI in the public sector were discussed, including work by the House of Lords and the Law Commission, research and civil society organisations. Yet participants identified a lack of coherence and standardisation regarding the implementation of data analytics and data sharing in the public sector and referred to a 'marketplace of data ethics frameworks'.

The workshop provided a snapshot of many of the tensions that exist within the wider discourse surrounding issues of technology and the public sector. Several controversies addressed the heart of relationships and interactions between the state, the general public, and private enterprise. The questions raised in this workshop have informed the research of the Data Scores as Governance project.

Journalist workshop (21 September 2018, Cardiff)

The workshop 'Investigating and reporting on government uses of algorithms and data' brought together national and regional journalists within the UK, data journalism educators and civil society actors, at Cardiff University's School of Journalism, Media and Culture on 21 September 2018. The aim of the workshop was to discuss and advance journalistic reporting on uses of data analytics and algorithmic decision-making in local and central government. To that end, the investigator team presented the Data Scores Investigation Tool (www.data-scores.org) to key users.

The event started with a keynote speech by Nick Diakopoulos (Northwestern University) on the characteristics and challenges of the emerging 'algorithms beat' in newsrooms. This was followed by a demonstration of the new interactive tool created by the Data Justice Lab for investigating uses of data analytics in public services. The tool provides access to a wide range of documents and can be used by journalists, civil society and interested public to explore where and how data analytics are used by the public sector in the UK. Participants tested the tool and provided useful feedback that fed into its further development. The workshop closed with a discussion on how to develop an algorithms beat in journalism education. Key themes emerging from those discussions included the challenge of demonstrating the impact of data uses in the public sector and algorithmic decision-making on people's lives; the lack of resources within news organisations to carry out investigations in this area; and the limited understanding of processes and technologies in much news coverage of big data, particularly in the context of public services.

UN investigation workshop (6 November 2018, Cardiff)

The Data Justice Lab organised a workshop in connection with the visit from the UN Special Rapporteur for extreme poverty and human rights to the UK to investigate the effects of austerity. The workshop took place on 6 November 2018 at Cardiff University's School of Journalism, Media and Culture, with the goal of providing relevant input for the Special Rapporteur on the use of digital technologies in welfare provision and possibilities for enhancing citizen participation in the deployment of data systems in the public sector. It brought together frontline staff and representatives of civil society organisations who shared experiences and discussed challenges.

Issues that were discussed included the transformation of citizen-facing services from human-based interactions to interactions via online platforms; the automation of decision-making and other processes by central and local welfare authorities; and the increased collection of

data about welfare claimants and sharing of that data between various government departments. Participants raised concerns that a lack of physical access to the internet, as well as of the confidence and capability of engaging with official websites, remain a problem for many people. Digital welfare systems, such as Universal Credit in the UK, place a stronger individual responsibility on the claimant, which is problematic in this context. Issues of false data processing persist, according to participants, often with serious consequences for claimants. The increase in data sharing between public authorities and the combination of different datasets, without transparency or explanation, was seen as a major concern. According to one example, data collection intended for child protection was used for national security purposes. Further, it was noted, that data analytics often equate living in poverty with 'risk', with serious implications for those who are affected. The focus of data analytics on 'risks' rather than 'rights' and conditionality-based rather than enabling-based design were highlighted as underlying problems in digital social welfare provision.

A wider range of issues were discussed, together with suggestions for improvements and public participation. These provided a rich set of inputs for the Special Rapporteur's investigation.

Final Event (19 November, London)

The research on 'Data Scores as Governance' was concluded with an event in London on 19 November 2018 where research results were presented by the investigators and discussed by panellists from the public sector, academia, and civil society, as well as an audience with further representatives from these stakeholder groups.

The first panel included Jen Persson (Defend Digital Me), Simon Burall (Involve), Godfred Boahen (British Association for Social Workers) and Tom Fowler (Integrated Analytics Hub, Bristol). Based on the research findings, panellists discussed specific challenges of the use of data analytics in the public sector. One issue that was addressed concerned the complexity of data analytics systems in practice. Panellists noted that this presented challenges for frontline workers who may not properly understand the system and may use it incorrectly or be tempted to bypass it altogether. A lack of understanding, but also institutional pressures and a belief in the value of data, may lead to situations where data analytics severely constrain human decisions, even when data is only supposed to inform those decisions. Technology, it was argued, thus drives decision-making even when it is just designed to have a support function.

Questions were raised regarding the ability of algorithmic processing to capture the detailed, fine-grained understanding of cases that social workers and other case workers have to address. In addition, once a person is flagged in the system as a person of risk or with a particular vulnerability, this is likely to lead to a cycle of ever-more intervention. In this context, criteria for impact and success were discussed, and concerns were raised that such criteria often focus on the saving of resources, for example the need for fewer case workers, rather than the quality of service.

While the need for modernising public services was recognised, the discourse of technological modernisation, according to participants with experience in the sector, creates a need for constant 'innovation' and thus a practice of 'data for data's sake', which sometimes supersedes a careful evaluation of needs and risks. Legislation, as was noted, does not always capture the complexity of data analytics systems and may provide insufficient guidance. Further, the role of commercial providers of data and data analytics systems was discussed as those are used heavily by public services and may set their own success criteria, and combine datasets, outside of public scrutiny.

Just like in the earlier stakeholder workshop, public involvement was highlighted as a necessary factor for making data analytics accountable. This would require not just an increase in public knowledge and the willingness of local authorities but, as one participant noted, a broader transformation of governance structures.

The second panel focused on policy implications and the question of what should be done in response to the research results. Prof Lilian Edwards (Newcastle University) discussed several aspects of recent data protection policy, particularly the EU General Data Protection Regulation (GDPR), in relation to the concerns raised by both the research and the other commentators. The GDPR includes several relevant rules, such as the right to explanation of data processes; the right not to be subject to entirely automated decision-making; limits to the processing of sensitive data; and limits to data sharing. However, in practice, these notions have severe limitations as, e.g., solely automated systems (as referred to in GDPR) rarely exist, and the right to explanation requires significant knowledge by the citizen. So while there are some advances, data protection remains underdeveloped.

Edwards criticised a focus on transparency in much of the data-related policy debate, as transparency alone does not provide remedies, does not address the unequal power relations between citizens and the state, and focuses on individualistic responses. Similarly, she raised concerns regarding an emphasis on ethics in policy responses to data-related challenges and maintained that a rigorous legal and regulatory framework would be more helpful to address these challenges.

Javier Ruiz (Open Rights Group) concurred but also addressed the 'limits of the regulatory approach' and urged participants to tackle the problem of datafication, rather than to just fix the consequences. This would involve interrogating the role of prediction in governance more generally and to develop bottom-up approaches to data governance and related issues, such as trust.

The event concluded the research project but demonstrated that the debate on the uses of data analytics in the public sector is only beginning.

Section VII:
Discussion

Discussion

In this section we outline a few key discussion points and policy implications that we can draw from our research into the implementation and uses of data analytics in public services, and the prevalence of data-driven scoring systems in decision-making, in particular. It is clear from our research that there is a need for a more nuanced and contextual debate about the uses of data across the public sector in the UK. We have found developments to be significantly distinct in different local authorities, with no standard procedures in place for how data systems are implemented, discussed and audited. We also found that uses of data systems are approached very differently, with some data-sharing leading to the creation of individual risk scoring, whilst in other contexts this is not practiced and databases serve predominantly as verification tools or population level analytics. This indicates that whilst it is broadly accepted that health and social service planning requires data and analytics, we do not have a shared understanding amongst local authorities as to what it is appropriate to do with such technologies.

Increased data sharing is a prominent trend across our case studies, with data warehouses serving to integrate data from different parts of councils and authorities to, in some cases, enable population level analytics and, in other cases, make it easier for frontline staff to share information and provide single citizen and household views. A recurring theme in the rationale for implementing data systems is the context of austerity, with managers and developers often responding to significant cuts by trying to use data to better target resources. This speaks to the contextual duality of data-driven technologies as one of data-rich and resource-poor contexts (Mcquillan, 2018a). Prominent uses of data scoring and risk assessment exist in areas such as child welfare and policing where vulnerability and risk are calculated through the combination of extensive data-sets that identify characteristics and behaviours of previous victims and offenders in order to flag individuals with similar characteristics. These scores and reports are provided to frontline workers as intelligence to help indicate who might need further attention. Importantly, our research across sectors found an emphasis on professional judgement and claims that no decisions are made on the basis of data-driven scores alone. This is important as a key concern expressed in debates about uses of data is the often de-contextualised nature of information. This issue was recognised by many of the people we interviewed and is frequently pertinent to systems adopted by frontline staff. However, our workshop discussions also point to the challenges of upholding genuine professional judgment and discretion in relation to increased use of data analytics. Within a broader context of deskilling and resource limitations in the public sector, the results of data analytics may significantly constrain and guide decision-making.

Our research found that auditing for accuracy levels and user activity of data systems is widespread, but more comprehensive auditing mechanisms vary greatly. Some local authorities do regular audits for data quality, which was noted by practitioners as a key challenge with using data systems in public services, and in some instances councils and police authorities have carried out privacy impact assessments and some citizen consultations. Generally, however, citizen participation and possibilities for intervention into how and why data systems are implemented remain elusive, as does any assessment of the impact of

interventions taken on the basis of data scores. We saw some moves to address this in some of our case studies, such as Kent's plans to engage more public feedback through public advisory and stakeholder groups as well as their current practice of stakeholder approval boards. The need for more public engagement remains a key concern amongst stakeholder groups in civil society who pointed to a fundamental lack of transparency regarding data developments in public services. As a wide range of research has shown, public knowledge about data analytics is low, and the details of data collection and analysis remain obscure to large parts of the population, often leading to frustration and resignation regarding the workings of data-based systems (Turow et al., 2015; Dencik & Cable, 2017). The 'black box' character of algorithmic governance has been highlighted repeatedly (Pasquale, 2015; O'Neil, 2016), and efforts have emerged in both academia and civil society to better understand algorithms and advance 'algorithmic accountability' (Diakopoulos, 2014). At a basic level, there have been calls for a list of where and when data systems are implemented in public services and what kind of auditing is carried out, in order to generate more public debate and understanding about uses of data. The recent EU General Data Protection Regulation (GDPR) assigns citizens a right to explanation of data processes and to challenge their outcomes, and it expands rules for consent to one's data collection.

However, while these approaches constitute important pre-conditions for responsible data use, they cannot, as our research shows, address all its challenges. Transparent processes of data analytics are still situated within unequal power relations between state institutions and the individual citizen. Knowledge by the citizen does not necessarily lead to effective remedy, nor does it address the responsibility of the state for careful and adequate treatment of individual citizens. Robust regulation is essential to guide the use of data analytics and broader institutional changes may be necessary as citizens are increasingly assessed and serviced according to data. This is the case, not least, with the growing role of private companies in the development and supply of data processing systems in the public sector. Although some local authorities have sought to develop these systems in-house, companies like Xantura, Callcredit and Capita are increasingly contracted to provide predictive analytics and risk assessments. We have seen how information about these systems, e.g. notifying people that their data is being used for these purposes, is circumvented due to concerns about it compromising commercial interests or the functioning of the system. These systems, as noted above, are often introduced as a means for local authorities to save money. What promises about savings are being made and how might the desire to cut costs influence the kinds of services offered? Given the ongoing issues with data quality and predictive accuracy, more transparent discussions about data system limits are required. With data processing systems affecting an ever-growing area of the public sector, effective mechanisms will be necessary to advance public participation in the implementation of such systems. This may involve consultations, citizen audits, or other forms of public involvement. Due to the current lack of public knowledge and the ineffectiveness of many public consultations, new directions may need to be explored to involve the citizenry.

This is particularly pressing as our research indicates a significant disparity between practitioners' and stakeholder groups' perspectives of the nature of challenges that emerge from uses of data analytics in public services. Whilst practitioners identify predominantly

technical challenges, such as the issue of data quality, along with cultural challenges within the organisation in terms of adoption, civil society groups point to a number of social and political challenges. These include, for example, questions of bias and discrimination as well as issues of stereotyping and stigmatisation. Here, there is a particular concern with marginalised groups and resource-poor families who might be especially impacted by risk identification, and once labelled as such, feel targeted. Criticisms of uses of predictive analytics systems in other countries have raised concerns about how systems like this, which disproportionately draw on and use data about people who make use of social services, are biased through the over-representation of a particular part of the population. The variables being used can in practice be proxies for poverty. For example, in other countries researchers have found the length of time someone was on benefits as a variable influencing risk assessments. These different perspectives indicate that debates are siloed and there is a lack of communication between those working with data systems and those who engage with impacted communities. Finding avenues to connect such debates is essential for a better understanding of how to balance efforts to use data to provide better services and address population needs and concerns over data collection, including an understanding of situations in which data may not be the solution to a problem.

Frameworks for data ethics have been developed by a range of actors, from scholars[141] to the UK government[142], and the new Centre for Data Ethics and Innovation[143] may institutionalise data ethics in a governmental context. Data ethics can offer important advances in the responsible treatment of data and can, specifically, complement necessary guidelines for services and institutions dealing with personal data. However, like the principle of transparency, it has significant limitations. Without being accompanied by a robust regulatory framework, data ethics risks transforming the protection of citizen rights into a self-guided act by public and private sector entities that is either voluntary or negotiated between those stakeholders. That alone would be an insufficient framework for a form of governance that deeply affects the wider citizenry. Legislative and regulatory rules for, and restrictions of, the collection and analysis of citizens' data are therefore essential. The GDPR addresses several relevant areas, including the right not to be subject to entirely automated decision-making; limits to the processing of sensitive data, such as "data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs [etc.]"; and limits to data sharing between commercial entities. Many elements of the GDPR have been controversial and are regarded as insufficient (e.g., Edwards & Veale, 2017; Wachter et al., 2017) but it has addressed some of the gaps in the regulation of data analytics and points to necessary avenues for future regulation.

A key controversy regarding the use of data analytics by public services concerns the aggregation of a wide variety of data about people. For public services, data sharing across different departments and the inclusion of a broad range of data can be useful. As our case studies have shown, some local authorities strive towards an 'integrated view of the citizen'

[141] https://www.oii.ox.ac.uk/news/releases/what-is-data-ethics/

[142] https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework

[143] https://www.gov.uk/government/groups/centre-for-data-ethics-and-innovation-cdei

and some commercial data-processing agencies combine demographic, transactional, social, and other types of data towards comprehensive scores and categories. For civil society groups, however, and for many service users and citizens, this wide use of data beyond its intended purpose constitutes a core problem of current implementations of data analytics. They argue for a strict purpose limitation of data collection and analysis - a principle supported by the GDPR which states that data must only be "collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes" (Art. 5). Stricter rules for the purpose limitation of data therefore address questions of both citizen rights and trust in the data practices of institutions, including government, and deserve careful attention.

Engaging with civil society concerns and assessments of the implications for impacted communities are especially pertinent as there is an underlying assumption in the implementation of data systems in public services that information will lead to action. However, without comprehensive evaluation of how these new data arrangements are, or are not, affecting action, engagement and resources, these claims remain unproven. Prominent assumptions exist, for example, that frontline staff are able to respond to risk, with little evidence provided about changes in resource allocation. There are also assumptions that early intervention and pre-emptive measures are inherently good, with little assessment of how being labelled may lead to unintentional harmful consequences. In particular, experiences amongst service-users and communities point to the need to engage more comprehensively with the way data systems relate to different forms of actions that might lead to a range of harms and feelings of being targeted. This requires a re-evaluation of how authorities and the state might be perceived as not necessarily benign, and that technologies are not necessarily neutral. Whilst harmful outcomes relating to data collection and use might not be intentional, such evaluations point to the need to consider how data has the potential to facilitate punitive measures under political cultures that deliberately target certain groups (e.g. hostility towards immigrants).

Moreover, the disentanglement of data systems, and the breakdown of the type of data and variables used to make assessments, point to the necessarily abstracted and reduced constitution of social knowledge and lived experiences that data analysis depends on in order to algorithmically process information. This is significant in several respects. In the case of Bristol's Integrated Data Analytics Hub, for example, it was recognised how data-driven risk assessments can only take account of risk factors such as school attendance, records of domestic abuse, etc. but cannot account for insulating 'positive' factors such as other types of social engagement or wider family networks that rely on contextual knowledge and unstructured information. Furthermore, whilst there are attempts to aggregate data to identify broader social issues that shape opportunities and challenges for families and individuals, household-level and individual-level data tends to personalise risk, privileging individualised responses over collective and structural responses. For example, we run the risk of measuring the impact of school absences but not the impact of school cuts. In other words, these systems, in their emphasis on correlation over causation, can individualise social problems by directing attention away from structural causes of social problems (Keddell, 2015; Andrejevic, 2017).

In the prominent application of data systems for the purposes of identifying and measuring risk, such as the widespread use of Risk Based Verification systems, we are also confronted with a general shift within public administration towards risk management as a new 'paradigm' of operations (Yeung, 2018). This is significant for the fundamental nature of state-citizen relations in an increasingly datafied society, suggesting that citizens are implicated in society not as participants or co-creators, but primarily as (potential) risks. In this context, the logic of capturing risk and granting authority to devices that calculate risk, can easily trump other forms of expertise or alternative forms of state-citizen relations that are informed by other values (Amoore, 2013). In our stakeholder workshops, a prominent discussion point concerned the extent to which frontline staff are being deskilled or disempowered as professionals with the growing use of data analytics. Whilst practitioners across councils and partner agencies emphasise the importance of professional judgement, experiences also indicate a push towards the rationalisation of the messiness of life and the limited parameters in which professional judgement can actually play out, particularly in a context of austerity and cuts to services. This suggests a need to engage more explicitly with the nature of knowledge created through data analytics, what it means to 'see' citizens through data, what lines of reasoning and argumentation are reinforced over others, and what the social and political consequences are (Redden, 2015; Hintz et al., 2018).

In general, our research points to a growing normalisation of data analytics and data-driven decision-making in public services. The fact that this is happening already suggests that data practices have become normalised before there has been a chance for broader public discussion (McQuillan, 2018b). There is a danger that the sheer fact that these systems are already in use will serve as a rationale for their continued existence and a means to foreclose debate. In carrying out this project, we hope to insert the possibilities for interventions and to force reflections on how to enhance a more democratic engagement with uses of data analytics in public services, and beyond.

References

Abi-Aad, G. (2016) Kent Public Health Observatory - Local datavores research workshop. 13th July 2016. Available at: https://www.youtube.com/watch?v=zgNowZ_UJAg

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

Amoore, L. & Piotukh, V. (2016) Introduction. In: L. Amoore and V. Piotukh (eds.), *Algorithmic Life: Calculative Devices in the Age of Big Data*. New York: Routledge, pp. 1–18.

Andrejevic, M. (2017) To pre-empt a thief. *International Journal of Communication*, 11, 879–96.

Angwin, J., Larson, J., Mattu, S. & Kirchner, L. (2016) Machine Bias. *Pro Publica*, 23rd May 2016. Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Aradou, C. & Blanke, T. (2015) The (Big) Data-security assemblage: Knowledge and critique. *Big Data & Society*, 2(2), 1–12.

Bass, T., Sutherland, E. & Symons, T. (2018) Reclaiming the Smart City: Personal data, trust and the new commons. Report published July 2018. Available at: https://www.nesta.org.uk/report/reclaiming-smart-city-personal-data-trust-and-new-commons/

Berry, D. (2011) The computational turn: Thinking about the digital humanities. *Culture Machine*, 12, 1–22. Available at: http://www.culturemachine.net/index.php/cm/article/viewDownloadInterstitial/440/470.

Big Brother Watch (2014) Care.data delay is not the end of the issue. Blog post, 19th February 2014. Available at: https://bigbrotherwatch.org.uk/2014/02/care-data-delay-end-issue/

Big Brother Watch (2018) A closer look at Experian big data and artificial intelligence in Durham Police. Blog post, 6th April 2018. Available at: https://bigbrotherwatch.org.uk/2018/04/a-closer-look-at-experian-big-data-and-artificial-intelligence-in-durham-police/

boyd, d. & Crawford, K. (2012) Critical questions for Big Data. *Information, Communication & Society*, 15(5), 662–79.

Brauneis, R. & Goodman, E. (2018) Algorithmic Transparency for the Smart City. *Yale Journal of Law & Technology* (20), 103-176

British Academy & The Royal Society (2017) Data management and use: Governance in the 21st century. Report issued June 2017. Available at: https://royalsociety.org/~/media/policy/projects/data-governance/data-management-governance.pdf

Burrows, R. & Gane, N. (2006) Geodemographics, Software and Class. *Sociology*, 40(5), 793-812

Cabinet Office (2012) Government Digital Strategy. November 2012. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/296336/Government_Digital_Stratetegy_-_November_2012.pdf

Cheney-Lippold, J. (2017) *We Are Data*. New York: New York University Press.

Chichester District Council (2017) Risk Based Verification Policy 2017. Available at: http://chichester.moderngov.co.uk/documents/s10306/Housing%20Benefit%20and%20Council%20Tax%20Reduction%20Risk%20Based%20Verification%20Policy%202018-2019%20-%20Appendix.pdf

Chin, J. & Wong, G. (2016) China's New Tool for Social Control: A Credit Rating for Everything. *The Wall Street Journal*, 28th November 2016. Available at: https://www.wsj.com/articles/chinas-new-tool-for-social-control-a-credit-rating-for-everything-1480351590

Coaffee, J. & Murakami Wood, D. (2006) Security is coming home: Rethinking scale and constructing resilience in the global urban response to terrorist risk. *International Relations*, 20(4), 503–17.

Corfield, G. (2018) Capita's UK military recruiting system has 'glitches' admits minister. *The Register*, 16th January 2018. Available at: https://www.theregister.co.uk/2018/01/16/capita_drs_has_glitches_defence_minister/

Crawford, K. (2013) The hidden biases in Big Data. *Harvard Business Review*, 1 April. Available at: https://hbr.org/2013/04/the-hidden-biases-in-big-data

Crawford, K. (2016). Know your terrorist credit score! Presentation at Re:publica, May, Berlin.

Cresswell, P., Holgate, M. & Smith, K. (2014) Under the bonnet: Mosaic data, methodology and build. Presentation delivered 1st April 2014. Available at: https://www.experian.co.uk/assets/marketing-services/presentations/mosaic-data-methodology-and-build.pdf

Crooks, R. (2017). Representationalism at work: Dashboards and data analytics in urban education. *Educational Media International*, 54(4), 289–303.

Dencik,L. & Cable, J. (2017). The Advent of Surveillance Realism: Public Opinion and Activist Responses to the Snowden Leaks. *International Journal of Communication*, 11(2017), 763-781.

Department for Work and Pensions (2011) Housing Benefit and Council Tax Benefit Circular HB/CTB S11/2011. 9th November 2011. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/633018/s11-2011.pdf

Diakopoulos, N. (2014) Algorithmic Accountability: On the Investigation of Black Boxes. Tow Center for Digital Journalism, 3rd December 2014. Available at: https://towcenter.org/research/algorithmic-accountability-on-the-investigation-of-black-boxes-2/

Dixon, P. & Gellman, R. (2014) The Scoring of America: How Secret Consumer Scores Threaten Your Privacy and Your Future. Report for the World Privacy Forum.

Durham County Council (2013) Ensuring that Children are Receiving a Suitable Education 2009. December 2013. Available at: http://www.durham.gov.uk/media/3062/Ensuring-that-children-are-Receiving-a-Suitable-Education/pdf/EnsuringThatChildrenAreReceivingASuitableEducation.pdf?m=636066154610030000

Echo, Basildon Canvey Southend (2018) Giggles as MP uses "Crapita" nickname in parliament. Basildon Canvey Southend Echo, 24th April 2018. Available at: http://www.echo-news.co.uk/news/16180590.Giggles_as_MP_uses__Crapita__nickname_in_parliament/

Edwards, L. & Veale, M. (2017) Slave to the algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for. *Duke Law and Technology Review*, 16(1), 18–84.

Eubanks, V. (2018) *Automating inequality: How high-tech tools profile, police, and punish the poor*. New York: St Martin's Press.

Experian (2006) Mosaic United Kingdom. Available at: https://www.prospectlists.co.uk/downloads/Mosaic_UK_2003_brochure.pdf

Experian (2007) Mosaic Global. Available at: https://www.experian.co.uk/assets/business-strategies/brochures/Mosaic_Global_factsheet[1].pdf

Experian (2010) Mosaic Global E-Handbook. Available at: http://www.appliedgeographic.com/AGS_2010%20web%20pdf%20files/Mosaic%20Global%20E-Handbook.pdf

Experian (2016) Mosaic Public Sector. Available at: https://www.experian.co.uk/assets/marketing-services/brochures/mosaic-ps-brochure.pdf

Experian (2018a) Mosaic: The consumer classification solution for consistent cross-channel marketing. Available at: https://www.experian.co.uk/assets/marketing-services/brochures/mosaic_uk_brochure.pdf

Experian (2018b) About Experian. Available at: https://www.experian.co.uk/about-us/index.html

Experian Global (2014) Inside Experian - The Full Film. 4th December 2014, Youtube. Available at: https://www.youtube.com/watch?v=YSxeXPD-p8g

Fink, K. (2017) Opening the government's black boxes: Freedom of information and algorithmic accountability. *Information, Communication & Society*. Epub ahead of print 20 May 2017. DOI: 10.1080/1369118X.2017.1330418.

Fullerton, J. (2018) China's 'social credit' system bans millions from travelling. *The Telegraph*, 24th March 2018. Available at: https://www.telegraph.co.uk/news/2018/03/24/chinas-social-credit-system-bans-millions-travelling/

Gangadharan, S.P., Eubanks, V. & Barocas, S. (eds.) (2015) Data and Discrimination: Collected Essays. Open Technology Institute, New America. Available at: http://newamerica.org/downloads/OTI-Data-an-Discrimination-FINAL-small.pdf

GDS (2018) GOV.UK Verify Overview. Available at: https://www.gov.uk/government/publications/introducing-govuk-verify/introducing-govuk-verify

Gillespie, T. (2014) The Relevance of Algorithms. In T. Gillespie, P. Boczkowski & K. Foot (eds.) *Media Technologies: Essays on Communication, Materiality and Society*. Cambridge: MIT Press.

Government Technology (2010) Government Technology, volume 9.7, published 28th October 2010. Available at: https://issuu.com/karlosullivan/docs/gt97magazine

Graham, J. (2017) London uses data to predict which children will be abused. *apolitical*, 18th September 2017. Available at: https://apolitical.co/solution_article/london-uses-data-predict-which-children-abuse/

Grieves, C. (2017) Joining the Dots Between Offline & Online Marketing. Uploaded to Youtube by Viant Technology. Available at: https://www.youtube.com/watch?v=XHOphdHCMhQ

Guardian, The (2018-) The Cambridge Analytica Files. Available at: https://www.theguardian.com/news/series/cambridge-analytica-files

Hall, M. & McCann, D. (2018) What's your score?: How discriminatory algorithms control access and opportunity. New Economics Foundation, 10th July 2018. Available at: https://neweconomics.org/2018/07/whats-your-score

Haringey Defend Council Housing (2017) Experian's Mosaic market segmentation system – in use by Lendlease. 27th August 2017. Available at: https://haringeydefendcouncilhousingblog.wordpress.com/2017/08/27/experians-mosaic-market-segmentation-system-in-use-by-lendlease/

Hern, A. (2018) Facebook among 30 organisations in UK political data inquiry. *The Guardian*, 5th April 2018. Available at: https://www.theguardian.com/technology/2018/apr/05/facebook-mark-zuckerberg-refuses-to-step-down-or-fire-staff-over-mistakes

Hill, R. (2018) Shiny new Capita boss to UK.gov: I know you are but what am I? *The Register*, 19th June 2018. Available at: https://www.theregister.co.uk/2018/06/19/shiny_happy_capita_new_boss_attempts_to_gloss_over_mps_concerns/

Hintz, A., Dencik, L. & Wahl-Jorgensen, K. (2018) *Digital Citizenship in a Datafied Society*. Cambridge: Polity Press.

HM Government (2012) Open data white paper. June 2012. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/78946/CM8353_acc.pdf

HM Government (2017) Government response to wider call for evidence from the House of Lords Committee on AI. Department for Digital, Culture, Media and Sport and Department for Business Energy and Industrial Strategy. Evidence dated 2nd November 2017. Available at: http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/artificial-intelligence-committee/artificial-intelligence/written/73250.html

Hussain, A. (2018) Revealed: how credit reference agencies get their facts wrong - and lose your data; Complaints and concerns increase over 'dodgy data'. *The Sunday Times*, 19th August 2018.

Hvistendahl, M. (2017) Inside China's Vast New Experiment in Social Ranking. *Wired*, 12 December 2017. Available at: https://www.wired.com/story/age-of-social-credit/

ICO (2018a) Democracy disrupted?: Personal information and political influence. 11th July 2018. Available at: https://ico.org.uk/media/action-weve-taken/2259369/democracy-disrupted-110718.pdf

ICO (2018b) Investigation into the use of data analytics in political campaigns. Available at: https://ico.org.uk/media/action-weve-taken/2259371/investigation-into-data-analytics-for-political-purposes-update.pdf

ICO (2018c) Emma's Diary Notice of Intent. 2nd July 2018. Available at: https://ico.org.uk/media/2259363/emmas-diary-noi-redacted.pdf

ITV News (2010) North Lincolnshire_Xantura VVPI Project. Uploaded to Youtube by user named chivvyneil, 18th November 2010. Available at: https://www.youtube.com/watch?v=Dq-I9E29Gsw

Jefferson, E. (2018) No, China isn't Black Mirror – social credit scores are more complex and sinister than that. *New Statesman*, 24th April 2018. Available at: https://www.newstatesman.com/world/asia/2018/04/no-china-isn-t-black-mirror-social-credit-scores-are-more-complex-and-sinister

Kelion, L. (2018) Emma's Diary faces fine for selling new mums' data to Labour. *BBC News*, 11th July 2018. Available at: https://www.bbc.co.uk/news/technology-44794635

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.

Kent County Council (2018a) Freedom of Information request response, reference number: 1185536. Response received 9th March 2018.

Kent County Council (2018b) Freedom of Information request response, reference number: 1189733. Response received 8th March 2018.

Khan, A. (2017) A Look At The Massive Data Centres Behind One of The World's Largest Credit Rating Firms. *Dazeinfo*, 28th June 2017. Available at: https://dazeinfo.com/2017/06/28/data-centres-the-worlds-largest-credit-rating-firms/

Kitchin, R. (2014) *The data revolution*. London: Sage.

Kitchin, R. (2017) Thinking critically about and researching algorithms. *Information, Communication & Society*, 20(1), 14–29.

Kitchin, R. & Lauriault, T.P. (2015) Towards critical data studies: Charting and unpacking data assemblages and their work. The Programmable City Working Paper 2. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2474112

Knapton, S. (2016) How the NHS got it so wrong with care.data. *The Telegraph*, 7th July 2016. Available at: https://www.telegraph.co.uk/science/2016/07/07/how-the-nhs-got-it-so-wrong-with-caredata/

Knight, W. (2017) The Dark Secret at the Heart of AI. *MIT Technology Review*, 11th April 2017. Available at: https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/

Krebs, B. (2013) Experian Sold Consumer Data to ID Theft Service. *KrebsOnSecurity*, 20th October 2013. Available at: https://krebsonsecurity.com/2013/10/experian-sold-consumer-data-to-id-theft-service/comment-page-2/

Lancashire County Council (2016) MOSAIC public sector 2016 analysis of Lancashire. Available at: http://www.lancashire.gov.uk/media/898791/mosaic-2016-research-monitor-v1.pdf

Lancashire County Council (2017) Mosaic analysis. Available at: http://www.lancashire.gov.uk/lancashire-insight/area-profiles/mosaic-analysis

Local Government Improvement and Development (2010) London Fire Brigade: Incident risk management toolkit. Available at: www.cfoa.org.uk/download/17664

Lv, A., & Luo, T. (2018) Asymmetrical Power Between Internet Giants and Users in China. *International Journal of Communication* 12(2018), 3877-3895.

Lyon, D. (2015) *Surveillance after Snowden*. Cambridge: Polity.

Malik, S. (2014) Civil servants deployed to help Capita clear PIP assessments backlog. *The Guardian*, 6th April 2014. Available at: https://www.theguardian.com/society/2014/apr/06/civil-servants-capita-pip-assessments-backlog

Malomo, F. & Sena, V. (2016) Data Intelligence for Local Government? Assessing the Benefits and Barriers to the Use of Big Data in the Public Sector. *Policy & Internet*, 9(1): 7-27.

Manning, M. & Toderas, A. (2017) The Benefits of Predictive Modelling in Councils. August 2017, published as part of the Catalyst Project, a collaboration between the University of Essex, Essex County Council, Suffolk County Council, and the Higher Education Funding Council for England. Available at: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=5&cad=rja&uact=8&ved=0ahUKEwjruNjW5YLcAhXR16QKHf52CvkQFghcMAQ&url=https%3A%2F%2Fwww.essex.ac.uk%2F-%2Fmedia%2Fdocuments%2Fresearch%2Fbenefits-of-predictive-modelling-in-councils.pdf&usg=AOvVaw2nkh9Ek6Wxg1_fF-6n_NxO

Massumi, B. (2015) *Ontopower: War, Powers, and the State of Perception*. Durham, NC: Duke University Press.

Mayer-Schönberger, V. & Cukier, K. (2013) *Big Data: A Revolution That Will Transform How We Live, Work and Think*. New York: John Murray.

Mazzucato, M. (2018) *The Entrepreneurial State: Debunking Public vs. Private Sector Myths*. Penguin.

McCann, D., Hall, M. & Warin, R. (2018) Controlled by calculations?: Power and accountability in the digital economy. New Economics Foundation, report published 29th June 2018. Available at: https://neweconomics.org/2018/06/controlled-by-calculations

McElhatton, N. (2004) Secrets of my Success: Richard Webber, Founder of geodemographics. *Campaign*, 31st October 2004. Available at: https://www.campaignlive.co.uk/article/secrets-success-richard-webber-founder-geodemographics/226747

McIntyre, N. & Pegg, D. (2018a) Councils use 377,000 people's data in efforts to predict child abuse. *The Guardian*, 16th September 2018. Available at: https://www.theguardian.com/society/2018/sep/16/councils-use-377000-peoples-data-in-efforts-to-predict-child-abuse

McIntyre, N. & Pegg, D. (2018b) Child abuse algorithms: from science fiction to cost-cutting reality. *The Guardian*, 16th September 2018. Available at: https://www.theguardian.com/society/2018/sep/16/child-abuse-algorithms-from-science-fiction-to-cost-cutting-reality

McIntyre, N. & Pegg, D. (2018c) Data on thousands of children used to predict risk of gang exploitation. *The Guardian*, 17th September 2018. Available at: https://www.theguardian.com/society/2018/sep/17/data-on-thousands-of-children-used-to-predict-risk-of-gang-exploitation

McQuillan, D. (2018a) Rethinking AI through the politics of 1968. *Opendemocracy,* 13 October. Available at: https://www.opendemocracy.net/digitaliberties/dan-mcquillan/rethinking-ai-through-politics-of-1968

McQuillan, D. (2018b) People's councils for ethical machine learning. *Social Media + Society*, 1-10. Available at: https://journals.sagepub.com/doi/full/10.1177/2056305118768303

Newcastle City Council (2017) Policy and Procedures for Children Missing or at Risk of Becoming Missing from Education (CME). Published October 2016 and updated April 2017. Available at: https://www.newcastle.gov.uk/sites/default/files/wwwfileroot/cme_policy_april_2017.pdf

Open Data Institute (2018) Using open data to deliver public services. Report published 2nd March 2018. Available at: https://theodi.org/article/using-open-data-for-public-services-report-2/

Omidyar Network & Upturn (2018) Public Scrutiny of Automated Decisions: Early Lessons and Emerging Methods. Report published 27th February 2018. Available at: https://www.omidyar.com/insights/public-scrutiny-automated-decisions-early-lessons-and-emerging-methods

O'Neil, C. (2016) *Weapons of math destruction: How big data increases inequality and threatens democracy*. Penguin.

ONS (2012) 2011 Census: Population Estimates for the United Kingdom, 27 March 2011. Available at: http://webarchive.nationalarchives.gov.uk/20160108132257/http://www.ons.gov.uk/ons/dcp171778_292378.pdf

Pagliery, J. (2014) Your personal information just isn't safe. *CNN Tech*, 28th July 2014. Available at: https://money.cnn.com/2014/07/25/technology/security/target-experian/

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, MA: Harvard University Press.

Pegg, D. & Hern, A. (2018) What triggered the ICO's political data inquiry? *The Guardian*, 11th July 2018. Available at: https://www.theguardian.com/uk-news/2018/jul/11/what-triggered-the-icos-political-data-inquiry

Redden, J. (2015). Big data as system of knowledge: investigating Canadian governance. In: G. Elmer, G. Langlois and J. Redden, J., eds., *Compromised Data: From Social Media to Big Data*, London: Bloomsbury.

Redden, J. & Brand, J. (2018) Data Harm Record. Available at: https://datajusticelab.org/data-harm-record/

Robson, S. (2005) Experian Data Centre, Nottingham. In: Bennett, D. (2005) *The Art of Precast Concrete.* Basel: Birkhäuser, pp. 114-115

Savage, M. & Burrows, R. (2007) The Coming Crisis of Empirical Sociology. *Sociology*, 41(5), 885-899

Science and Technology Committee, House of Commons (2018) Algorithms in decision-making. Report published 23rd May 2018. Available at: https://publications.parliament.uk/pa/cm201719/cmselect/cmsctech/351/351.pdf

Scott, K. (2018) Data for the Public Benefit: Balancing the risks and benefits of data sharing. Report published April 2018. Available at: https://www.involve.org.uk/resources/publications/project-reports/data-public-benefit

Select Committee on Artificial Intelligence, House of Lords (2018) AI in the UK: ready, willing and able? Rreport published 16th April 2018. Available at: https://www.parliament.uk/business/committees/committees-a-z/lords-select/ai-committee/news-parliament-2017/ai-report-published/

Symons, T. (2016a) Datavores of local government: Using data to make services more personalised, effective and efficient. Discussion paper published July 2016. Available at: https://media.nesta.org.uk/documents/local_datavores_discussion_paper-july-2016.pdf

Symons, T. (2016b) Wise council: Insights from the cutting edge of data-driven local government. Report published November 2016. Available at: https://www.nesta.org.uk/report/wise-council-insights-from-the-cutting-edge-of-data-driven-local-government/

Taylor, R. (2006) The rise and rise of Capita. *The Guardian*, 23rd March 2006. Available at: https://www.theguardian.com/business/2006/mar/23/partyfunding.society

Thielman, S. (2015) Experian hack exposes 15 million people's personal information. The Guardian, 2nd October 2015. Available at: https://www.theguardian.com/business/2015/oct/01/experian-hack-t-mobile-credit-checks-personal-information

Travis, A. (2014) Grayling awards Capita six-year UK electronic tagging contract. *The Guardian*, 15th July 2014. Available at: https://www.theguardian.com/business/2014/jul/15/capita-electronic-tagging-grayling-uk-contract-6-years-g4s-serco

TransUnion (2018) Callcredit becomes TransUnion as brand integration begins. *TransUnion Blog*, 12th July 2018. Available at: https://www.callcredit.co.uk/press-office/news/2018/07/callcredit-becomes-transunion

Tucker, P. (2016) Refugee or Terrorist? IBM thinks its software has the answer. *Defense One*. Available at: http://www.defenseone.com/technology/2016/01/refugee-or-terrorist-ibm-thinks-its-software-has-answer/125484/

Turow, J, Hennessy, M. & Draper, N. (2015).The Tradeoff Fallacy: How Marketers are Misrepresenting American Consumers and Opening Them Up to Exploitation. Report for Annenberg School of Communication. https://www.asc.upenn.edu/sites/default/files/TradeoffFallacy_1.pdf

Uprichard, E., Burrows, R. & Parker, S. (2009) Geodemographic code and the production of space. *Environment and Planning*, 41, 2823-2835.

Vagle, J.L. (2016) The history, means, and effects of structural surveillance. University of Pennsylvania Law School, Public Law Research Paper, 16-3.

Van Dijck, J. (2014) Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, 12(2), 197–208.

Wachter, S., Mittelstadt, B. & Floridi, L. (2017) Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *International Data Privacy Law*, 7(2), 76–99.

Warrell, H. (2015) Students under surveillance. *Financial Times*. Available at: https://www.ft.com/content/634624c6-312b-11e5-91ac-a5e17d9b4cff

Walker, P. J. (2016) Private firms earn £500m from disability benefit assessments. *The Guardian*, 27th December 2016. Available at: https://www.theguardian.com/society/2016/dec/27/private-firms-500m-governments-fit-to-work-scheme

Webber, R. (2007) Using names to segment customers by cultural, ethnic or religious origin. *Journal of Direct, Data and Digital Marketing Practice*, 8(3), 226-242

Webber, R. (2008) Names: a source of customer insight. Experian white paper. Available at: http://www.experian.nl/assets/documentatie/white-papers/names-a%20source-of-customer-insight-december-2008.pdf

Webber, R. & Burrows, R. (2018) The Predictive Postcode: The Geodemographic Classification of British Society. London: SAGE.

Yeung, K. (2018) Algorithmic government: Towards a New Public Analytics? Paper presented at ThinkBig, Windsor, 25 June.

Zuboff, S. (2015) Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30, 75–89.

Appendix 1 – Example targeted FOI request

The following is an example of the sort of targeted Freedom of Information request we sent when researching the case studies within this report.

Dear the London Borough of Camden,

We are submitting the following Freedom of Information Request. This request concerns the Camden Residents' Index and any related fraud detection systems that utilise the Camden Residents' Index. This is one of two related requests.

These fraud detection systems will relate (at least) to housing, council tax evasion, and school placements. Any additional areas of fraud protection where the Camden Residents' Index is utilised should also be included.

Specifically, we are requesting:

· Any briefing notes, reports or evaluations, reports or summaries to officials or others about the Camden Residents' Index and any related fraud detection systems that utilise the Camden Residents' Index.

· Any overviews about how the Camden Residents' Index and any related fraud detection systems that utilise the Camden Residents' Index work.

· Any data visualization outputs connected to the Camden Residents' Index and any related fraud detection systems that utilise the Camden Residents' Index.

· Any promotional material, presentation material, or educational material related to the Camden Residents' Index and any related fraud detection systems that utilise the Camden Residents' Index.

Also, please note we are not requesting datasets or any data about individuals.

Please let us know if any information requires clarifying.

We would like to receive electronic copies of this information.

All the very best,

Data Justice Lab // www.datajusticelab.org // datajusticelab@cardiff.ac.uk
School of Journalism, Media and Culture, Cardiff University

Appendix 2 – General FOI request

The general Freedom of Information request we sent to 403 UK local authorities during our second round of requests. Future publications will cover this aspect of our research in more detail.

Dear [Local Authority],

We are submitting the following Freedom of Information Request for documents relating to uses of data analytics, predictive analytics, or algorithmic automated systems used for risk assessment, scoring systems or automated decision making within the jurisdiction of your authority. This could include the use of these data systems in order to draw insights from large and integrated datasets, or to influence decisions about resources, funding or service delivery. This could include, but may not be limited to, uses of predictive analytics in child welfare, health care, adult social care, fraud detection, or policing.

Specifically, we are requesting:

· Any briefing notes, reports, evaluations, or summaries concerning uses of algorithmically driven data systems. Ideally this would include information about software, hardware, operations and types of data

· Any overviews about how system outputs are produced.

· Any training manuals or materials (e.g. slides, etc.) for staff about data collection, analysis, assessment and reporting as related to the use of the systems.

. Any impact assessments

· Any data visualisation outputs connected to the systems.

· Any promotional material, presentation material, or staff educational material related to the systems.

. Any contracts relating to the systems. We argue that the release of contracts does not compromise commercial sensitivity. The full argument is provided below.*

Also, please note we are not requesting datasets or any data about individuals.

Please let us know if any information requires clarifying.

We would like to receive electronic copies of this information.

All the very best,

Data Justice Lab

Please note that we are following the Information Commissioner's guidance on the Freedom of Information Act which specifies that requests can be made by organisations and there is no need for a named individual (pages 9-10, sections 38-39). Source: https://ico.org.uk/media/for-organisatio...

*Addressing the commercial sensitivity question:
a) Section 20 of the Department for Communities and Local Government's 'Local Government Transparency Code 2015', which encourages greater transparency in the public interest, notes:
'The Government has not seen any evidence that publishing details about contracts entered into by local authorities would prejudice procurement exercises or the interests of commercial organisations, or breach commercial confidentiality unless specific confidentiality clauses are included in contracts. Local authorities should expect to publish details of contracts newly entered into – commercial confidentiality should not, in itself, be a reason for local authorities to not follow the provisions of this Code'.
Source: https://assets.publishing.service.gov.uk... (page 9, section 20)

b) There is always the potential for a company in any area to act in bad faith, but this should not be a reason to deny access to information about how public money is spent.

c) Contract and tendering details are provided regularly by a range of government bodies in the public interest. For example, many police and fire services release contractual information through the Bluelight database (https://www.blpd.gov.uk/foi/foi.aspx).

d) There may be core aspects of the systems contracted that have competitive commercial value and we appreciate that such details will be redacted from any documents released. We argue that commercial interests will not be compromised by revealing other details that may be provided in contracts or related communications with contracted companies. Details that could be released include details about the existence of a system, it's purpose, the tasks being contracted, continued relationships, general descriptions, types of data being used, how this data is shared.

Appendix 3 – Sample practitioner interview questions

Theme 1: Details of system

1) What data is collected, analysed and shared (internally and externally)?

2) Where does it come from?

a. In what context? For what purpose?

3) How did you decide what data to include in the iBase database? Will more data be added? Are there any datasets that are missing that should be added in your view?

4) What were the challenges in constructing the database and how were they overcome?

5) What benefits have you seen?

6) Are risk assessments produced? Or any other forms of assessment or score?

Theme 2: Implementation of system

1) How was it decided to use data analytics?

2) Is the process audited? Is the impact assessed? (when?)

3) What have been key challenges in implementing the system?

4) What are the key benefits that come from using this system?

5) Do you receive any help from anyone else to help you make use of the data system?

a. Consulting with other local authorities?

b. Data Science Campus?

c. Private or third sector consultants?

Theme 3: Practices of system

1) How do you operationalize any risk assessments or insights gained from the system? Can you discuss examples?

2) To what extent does use of the database inform decision-making? What other factors are considered?

3) What are opportunities for contextualizing data?

4) Can you provide an example to detail how the system works in practice?

5) What happens once identifications/decisions have been made? What are the actions taken?

6) What are possibilities for practitioners to amending, reject or change assessments or other types of information once it is in the database?

7) Can families see the data that is held about them? What are opportunities for feedback from impacted communities? How is this considered?

8) What have been the lessons learned to date, through the development and use of this data system? Will its use change in the future, how?

Appendix 4 – Civil society sample interview questions

1. Could you outline the work and focuses of your organisation and your role within that work?

2. In your areas of focus, in what ways do you see data or data analytics influencing decision-making?

3. What are the opportunities that uses of data analytics in public services might provide?

4. What are your concerns about the use of data analytics in the public sector?

5. How do you understand harms produced by data systems? Are you seeing people being negatively affected by uses of data analytics in your work?

6. How might risks relating to the use of data analytics be mitigated?

7. What potential do you see for a) data ethics and b) regulation to address concerns?

8. What is missing in efforts to mitigate data harms?

9. Are you actively engaging with this debate? If yes, who are you working with?

10. Anything we haven't touched upon that you would like to speak to?

Appendix 5 - Systems discovered through FOI requests

Below is a complete list of the systems or related processes mentioned within our Successful responses, alongside the name of the Local Authority claiming to use the system. We have also included free text responses where a system has been referred to but given a name. For the exploratory requests, we have also included a link to the request on WhatDoTheyKnow. The list is alphabetised by Local Authority name. The list may of be interest to anyone wishing to research data analytics systems at the local or national level.

Local Authority	FOI Link	System Name
Ards and North Down Borough Council	https://www.wha	Google Analytics
Argyll and Bute Council	https://www.wha	Nuance Voice Automation Service
Argyll and Bute Council	https://www.wha	Sentiment Metrics social media sentiment analysis
Birmingham City Council	https://www.wha	"a suite of reporting tools to extract data and present for scorecards, reports and data extracts"
Birmingham City Council	https://www.wha	"cost calculations for financial purposes"
Birmingham City Council	https://www.wha	Business Objects
Birmingham City Council	https://www.wha	Civica Management System
Birmingham City Council	https://www.wha	Crystal Reports
Birmingham City Council	https://www.wha	IDEA
Birmingham City Council	https://www.wha	SmartExporter
Birmingham City Council	https://www.wha	Tableau
Blaby District Council	https://www.wha	Mosaic (Experian)
Blackburn and Darwen Borough Council	https://www.wha	Traffic signalling, Siemens & TRL Software
Bournemouth Borough Council	https://www.wha	"statistical reports on usage of online resources"
Bournemouth Borough Council	https://www.wha	"various traffic surveying and counting methods to understand traffic flows"
Bournemouth Borough Council	https://www.wha	AccsMap
Bournemouth Borough Council	https://www.wha	Arcady
Bournemouth Borough Council	https://www.wha	Capita Alto Library Management System
Bournemouth Borough Council	https://www.wha	GIS
Bournemouth Borough Council	https://www.wha	Google Analytics
Bournemouth Borough Council	https://www.wha	Google Tag Manager
Bournemouth Borough Council	https://www.wha	GovDelivery
Bournemouth Borough Council	https://www.wha	Hotjar SiteImprove
Bournemouth Borough Council	https://www.wha	LINSIG
Bournemouth Borough Council	https://www.wha	Mova
Bournemouth Borough Council	https://www.wha	Paramics
Bournemouth Borough Council	https://www.wha	PEARS
Bournemouth Borough Council	https://www.wha	SATURN
Bournemouth Borough Council	https://www.wha	Scoot
Bournemouth Borough Council	https://www.wha	SocialSignIn
Bournemouth Borough Council	https://www.wha	Stratos
Bournemouth Borough Council	https://www.wha	Tableau
Bournemouth Borough Council	https://www.wha	TEMPRO
Bournemouth Borough Council	https://www.wha	Traffic Master
Bournemouth Borough Council	https://www.wha	TUBA
Bournemouth Borough Council	https://www.wha	VISSIM
Bournemouth Borough Council	https://www.wha	WM Design GIS
Bournemouth Borough Council	https://www.wha	WRATE life cycle environmental impact assessment software
Bradford City Council	https://www.wha	CapitaONE
Bradford City Council	https://www.wha	LCS (including Early Help)
Bradford City Council	https://www.wha	SystemOne - Adult
Braintree District Council	https://www.wha	"Data Matching system used by all Essex authorities"
Braintree District Council	https://www.wha	National Fraud Initiative
Breckland District Council	https://www.wha	ACORN
Brent Borough Council	https://www.wha	National Fraud Initiative
Brent Borough Council	https://www.wha	Risk Based Verification
Brighton and Hove City Council	https://www.wha	"carrying out data-matching exercises, including maintaining a cross-reference (Customer Index)"
Brighton and Hove City Council	https://www.wha	"converting geographic datasets into mapping layers"
Brighton and Hove City Council	https://www.wha	"delivering projects involving advanced analytics (e.g. prediction and modelling)"
Brighton and Hove City Council	https://www.wha	"ensuring that relevant address datasets are maintained"
Brighton and Hove City Council	https://www.wha	"maintenance of semantic layers to simplify reporting"
Brighton and Hove City Council	https://www.wha	"managing online GIS portals"
Brighton and Hove City Council	https://www.wha	"managing the Street Numbering & Naming service"
Brighton and Hove City Council	https://www.wha	"traditional reporting bureau"
Brighton and Hove City Council	https://www.wha	ArcGIS
Brighton and Hove City Council	https://www.wha	Business Objects
Brighton and Hove City Council	https://www.wha	Excel (Microsoft)
Brighton and Hove City Council	https://www.wha	FME
Brighton and Hove City Council	https://www.wha	Gephi
Brighton and Hove City Council	https://www.wha	Infogram

Local Authority	FOI Link	System Name
Brighton and Hove City Council	https://www.wha	KNIME
Brighton and Hove City Council	https://www.wha	Lumira
Brighton and Hove City Council	https://www.wha	Neo4j
Brighton and Hove City Council	https://www.wha	Power BI (Microsoft)
Brighton and Hove City Council	https://www.wha	Predictive Analytics (SAP)
Brighton and Hove City Council	https://www.wha	R & Python [?in-house system developed]
Brighton and Hove City Council	https://www.wha	Spark
Brighton and Hove City Council	https://www.wha	Tableau
Brighton and Hove City Council	https://www.wha	Tableau Prep
Bristol City Council	[Not via WDTK]	Qlik
Bristol City Council	[Not via WDTK]	Think Family
Carlisle City Council	https://www.wha	Housing Benefit System (Capita)
Carlisle City Council	https://www.wha	Risk Based Verification (Xantura)
Ceredigion Council	https://www.wha	Daffodil
Ceredigion Council	https://www.wha	LDP
Ceredigion Council	https://www.wha	POPGROUP (Edge Analytics)
Charnwood Borough Council	https://www.wha	Abritas Shortlisting
Charnwood Borough Council	https://www.wha	QL Rent Arrears Progression
Chesterfield Borough Council	https://www.wha	"The 'Guage' system"
Chiltern District Council	https://www.wha	Risk Based Verification
City of York Council	https://www.wha	"a centralised performance management system (aggregated statistics on operational business areas / trends) which we use to provide an evidence base for decisions about resources, funding or service delivery"
City of York Council	https://www.wha	"a master data management system to improve data quality throughout the organisation"
Copeland Borough Council	https://www.wha	GIS (Geographical Information Systems)
Crawley Borough Council	https://www.wha	Business Objects
Crawley Borough Council	https://www.wha	IDEA
Croydon Borough Council	https://www.wha	Business Objects
Croydon Borough Council	https://www.wha	Croydon Observatory
Croydon Borough Council	https://www.wha	Excel (Microsoft)
Croydon Borough Council	https://www.wha	Power BI (Microsoft)
Dacorum Borough Council	https://www.wha	Risk Based Verification CiC (CallCredit)
Dacorum Borough Council	https://www.wha	Risk Based Verification Claims (CallCredit)
Dartford Borough Council	https://www.wha	Taxi Driver English Test (Versant)
Derbyshire Dales District Council	https://www.wha	HSE approved spreadsheet
Derbyshire Dales District Council	https://www.wha	M3PP -- "calculates risk scores ... for Food Safety, Health & Safety, Environmental Permitting and Private Sector Housing Environmental Health"
Dudley Council	https://www.wha	"a core dashboard of measures that is updated on a monthly basis but adapted to theneedds of different groups"
Dudley Council	https://www.wha	"Analysis of activity and potential future demand for Complex and Inclusion serice"
Dudley Council	https://www.wha	"Analysis of methods for identifying people at risk of loneliness and isolation - to be presented at Public Health England annal conference"
Dudley Council	https://www.wha	"Attainment data analysers"
Dudley Council	https://www.wha	"Business modelling of processes to inform reporting requirements, including: Community reablement, Assessment and support plans, Delayed transfers of care, transition and communication passports, Reviews, Waiting list process"
Dudley Council	https://www.wha	"Commissioning support through provision of data on Reablement"
Dudley Council	https://www.wha	"Connextions data reporting and administration of client caseload information system"
Dudley Council	https://www.wha	"Data analysis of outcomes for PVI and other early years settings"
Dudley Council	https://www.wha	"Education and school effectiveness dashboards including information such as attendance, exclusions, elective home education, school Ofsted inspections and SEND"
Dudley Council	https://www.wha	"Health protection and surveillance: The team analyses data relating to air quality, immunisations, winter surveillance and infection control, in order to guide these preventative programmes."
Dudley Council	https://www.wha	"Movelling impacts of SEN provision options, linked to School Place Planning process"
Dudley Council	https://www.wha	"Performance reporting: Implemented more detailed analysis of complaints and recording of data"
Dudley Council	https://www.wha	"Population Hub, as part of thhe Office for Data Analytics"
Dudley Council	https://www.wha	"Population profiling: The team conduct a range of analysis of the local population, monitoring trends in birth and death rates, as well as providing information on the diverse make-up of our communities."
Dudley Council	https://www.wha	"School problem profiles, providing indications around vulnerability"
Dudley Council	https://www.wha	"The schools data package is a traded service, including advice and support to schools"
Dudley Council	https://www.wha	"The team conduct analysis to support a range of lifestyle support services that are delivered across the borough"
Dudley Council	https://www.wha	Adoption Leadership Board (ALB) quarterly dataset
Dudley Council	https://www.wha	Adult Social Care Finance Return
Dudley Council	https://www.wha	Business Objects (SAP)
Dudley Council	https://www.wha	Children in Need (CIN) Census
Dudley Council	https://www.wha	Children Looked After (903 return)
Dudley Council	https://www.wha	Cost Savings Calculator (CSC) [Troubled Families]
Dudley Council	https://www.wha	Deprivation of Liberty Safeguards (DoLS)
Dudley Council	https://www.wha	Disability Service [dashboard]
Dudley Council	https://www.wha	Family Progress Data (FPD) [Troubled Families]

Local Authority	FOI Link	System Name
Dudley Council	https://www.wha	National Fraud Initiative
Dudley Council	https://www.wha	National Impact Study (NIS) [Troubled Families]
Dudley Council	https://www.wha	Ofsted Fostering Dataset
Dudley Council	https://www.wha	Power BI (Microsoft)
Dudley Council	https://www.wha	R
Dudley Council	https://www.wha	Safeguarding Adults Collection
Dudley Council	https://www.wha	Short and Long Term (SALT) Return
Dudley Council	https://www.wha	Single Person DIscount Review (TransUnion/CallCredit) -- used by external service provider, Civica
Dudley Council	https://www.wha	SSRS
Dudley Council	https://www.wha	Vulnerable Young Person's Tool [Troubled Families]
Ealing Borough Council	https://www.wha	"Experian provide Business solutions throughout the council services such as payment information/data matching"
Ealing Borough Council	https://www.wha	London Counter Fraud Hub Initiative
Ealing Borough Council	https://www.wha	Risk Based Verification (Coactiva)
East Hampshire District Council	https://www.wha	Dynamics (Microsoft)
East Hampshire District Council	https://www.wha	Experian Public Sector profiler [possibly a reference to Experian's Mosaic]
East Hampshire District Council	https://www.wha	GIS application
East Riding of Yorkshire Council	https://www.wha	Risk Based Verification (Xantura & Northgate PS Ltd)
Elmbridge Borough Council	https://www.wha	NFI [maybe referring to the National Fraud Initiative]
Erewash Borough Council	https://www.wha	Risk Based Verification
Exeter City Council	https://www.wha	Exeter Data Mill
Falkirk Council	https://www.wha	Living Well Falkirk / Falkirk Health & Social Care Partnership (ADL Smartcare)
Folkestone & Hythe District Council	https://www.wha	Risk Based Verification (Xantura)
Fylde Borough Council	https://www.wha	Risk Based Verification (TransUnion/Callcredit)
Glasgow City Council	https://www.wha	"using data analytics in an exploratory capacity, on ad hoc projects"
Greater Manchester Combined Authority	[Not via WDTK]	"a number of CadCorp tools"
Greater Manchester Combined Authority	[Not via WDTK]	"a Workload Modelling tool"
Greater Manchester Combined Authority	[Not via WDTK]	"bespoke Risk SharePoint site"
Greater Manchester Combined Authority	[Not via WDTK]	"In-house Performance Portal and Data Warehouse"
Greater Manchester Combined Authority	[Not via WDTK]	CAMMS Risk, integrated planning system
Greater Manchester Combined Authority	[Not via WDTK]	Child Passport
Greater Manchester Combined Authority	[Not via WDTK]	GIS
Halton Borough Council	https://www.wha	Business Objects (SAP)
Halton Borough Council	https://www.wha	CareFirst
Halton Borough Council	https://www.wha	Excel (Microsoft)
Halton Borough Council	https://www.wha	WEBI
Harrogate Borough Council	https://www.wha	Excel (Microsoft)
Harrogate Borough Council	https://www.wha	Microsoft Access database
Harrogate Borough Council	https://www.wha	REM (Regional Econometric Model)
Havant Borough Council	https://www.wha	Dynamics (Microsoft)
Havant Borough Council	https://www.wha	Experian Public Sector profiler [possibly a reference to Experian's Mosaic]
Havant Borough Council	https://www.wha	GIS application
Hertfordshire County Council	https://www.wha	"We also profile to identify people who might be interested in supporting our work in specialised areas (for example, as foster carers or adoptive parents)."
Hertfordshire County Council	https://www.wha	"We analyse our existing service data to make sure we can provide the services needed in the future"
Hertfordshire County Council	https://www.wha	Mosaic (Experian)
Highland Council	https://www.wha	Mendel (TotalMobile)
Hull City Council	https://www.wha	Risk Based Verification
Huntingdonshire District Council	https://www.wha	"There are some examples of algorithm and/or automation including: form validation and data entry"
Huntingdonshire District Council	https://www.wha	Northgate case management system
Huntingdonshire District Council	https://www.wha	Risk Based Verification (CallCredit)
Inverclyde Council	https://www.wha	Eligibility Criteria (EC) (Health & Social Care Partnership)
Inverclyde Council	https://www.wha	Indicator of Relative Need (IoRN) (Health & Social Care Partnership)
Inverclyde Council	https://www.wha	Scottish Patients at Risk of Readmission and Admission (SPARRA) (Health & Social Care Partnership)
Ipswich Borough Council	https://www.wha	Risk Based Verification (CallCredit)
Islington Borough Council	https://www.wha	FACE toolset (Imosphere)
Islington Borough Council	https://www.wha	Holistix (Quality Education Systems (QES)) [data matching tool]
Kent County Council	[Not via WDTK]	ACORN (CACI)
Kent County Council	[Not via WDTK]	Electronic Frailty Index
Kent County Council	[Not via WDTK]	Kent Integrated Dataset (KID)
Kent County Council	[Not via WDTK]	Kings Fund Combined Predictive Model
Kent County Council	[Not via WDTK]	Mosaic (Experian)

Local Authority	FOI Link	System Name
Kent County Council	[Not via WDTK]	SPSS
Kent County Council	[Not via WDTK]	STATA
Knowsley Metropolitan Borough Council	https://www.wha	"The Council manually discovers, interprets and communicates meaningful patterns in the data it collects, processes and stores."
Lancashire County Council	[Not via WDTK]	InstantAtlas (Geowise)
Lancashire County Council	[Not via WDTK]	Lancashire Insight
Lancashire County Council	[Not via WDTK]	Power BI (Microsoft) [for their "Neighourbood Intelligence Tool"]
Lancashire County Council	[Not via WDTK]	Umbraco content management system
Leeds City Council	https://www.wha	"Schools Admission process for allocating pupil places"
Leeds City Council	https://www.wha	"Youth Justice system, Asset Plus, has a process for managing contacts"
Leeds City Council	https://www.wha	FFT Aspire
Liverpool City Council	https://www.wha	"we share our data ... with the CCG in a psuedonymised format at client level to facilitate primary care risk stratification based on the "Welsh model" for large scale health risk identification"
London Borough of Hackney	[Not via WDTK]	"London Borough of Hackney carries out a range of data analytics projects and systems"
London Borough of Southwark	[Not via WDTK]	Local Care Record health records system
London Borough of Southwark	[Not via WDTK]	Student council tax discount review, with National Fraud Authority Initiative & Fujitsu
Luton Borough Council	https://www.wha	Capita tool used "in Revenues & Benefits" [perhaps this is Risk Based Verification]
Luton Borough Council	https://www.wha	FACE tool
Maidstone Borough Council	https://www.wha	Power BI (Microsoft), used by Parking Services
Manchester Clinical Commissioning Group	[Not via WDTK]	CareCentric
Manchester Clinical Commissioning Group	[Not via WDTK]	Combined Predictive Model
Manchester Clinical Commissioning Group	[Not via WDTK]	EMIS/GP
Medway Council	https://www.wha	Academy, used by Housing
Medway Council	https://www.wha	Acolaid, used by Licensing
Medway Council	https://www.wha	Clarity, used by Leisure
Medway Council	https://www.wha	Corporate Service Desk (TopDesk)
Medway Council	https://www.wha	Council Boards (ModernGov)
Medway Council	https://www.wha	Customer Contact (Lagan)
Medway Council	https://www.wha	Diabetes Prevention programme
Medway Council	https://www.wha	EDRMS - IDOX
Medway Council	https://www.wha	Equitrac, used by Priting
Medway Council	https://www.wha	Eros, used by Democratic Services
Medway Council	https://www.wha	Frameworki, used by Social Care
Medway Council	https://www.wha	Health weight programmes
Medway Council	https://www.wha	Highways, Streetworks, Waste, Greenspaces - Confirm - Asset Register
Medway Council	https://www.wha	ICAM, used by Library Public Access PCs
Medway Council	https://www.wha	IKEN, used by Legal
Medway Council	https://www.wha	Integra, used by Finance
Medway Council	https://www.wha	iworld, used by Revenues and Benefits
Medway Council	https://www.wha	IYSS, used by Youth Offending Team
Medway Council	https://www.wha	Jadu
Medway Council	https://www.wha	National Child Measurement Programme
Medway Council	https://www.wha	NHS Health Checks programme "which involves use of a nationally approved algorithm to calculate a person's risk of cardiovascular event in the next 10 years"
Medway Council	https://www.wha	Qmatic
Medway Council	https://www.wha	ResourceLink, used by Human Resources
Medway Council	https://www.wha	RSS2000, used by Registrars
Medway Council	https://www.wha	SAM
Medway Council	https://www.wha	Spektrix, used by Theatres
Medway Council	https://www.wha	Spudus, used by Libraries
Medway Council	https://www.wha	Tarranto, used by Parking
Medway Council	https://www.wha	Uniform, used by Planning
Melton Borough Council	https://www.wha	Mosaic
Milton Keynes Council	https://www.wha	NHS Health Check software
Neath Port Talbot County Borough Council	https://www.wha	IDEA Data Analysis (CaseWare)
New Forest District Council	https://www.wha	DRS Cyclical Servicing Module (Kirona)
New Forest District Council	https://www.wha	DRS Scheduler (Kirona)
Newcastle under Lyme Borough Council	https://www.wha	Geographic Information Systems
North Hertfordshire District Council	https://www.wha	Pentana system
North Somerset Council	https://www.wha	"In addition to Business Intelligence, we also have a small number of systems that 'profile' our service users."
North Somerset Council	https://www.wha	Business Intelligence - Consultation mailings (Inovem Ltd)
North Somerset Council	https://www.wha	High Impact Families Access database
North Somerset Council	https://www.wha	ICT and Transformation - Cookies (Agilisys Ltd)

Local Authority	FOI Link	System Name
North Somerset Council	https://www.wha	Revenues and Benefits - Benefits service (Liberata)
North Tyneside Metropolitan Borough Council	https://www.wha	Alteryx
North Tyneside Metropolitan Borough Council	https://www.wha	Tableau
Northampton Borough Council	https://www.wha	Azure Vision APIs
Northampton Borough Council	https://www.wha	Google Cloud Vision API
Northampton Borough Council	https://www.wha	Rekognition (Amazon)
Northamptonshire County Council	https://www.wha	Abacus Social Care financial assessment system
Northamptonshire County Council	https://www.wha	AssetPlus
Northamptonshire County Council	https://www.wha	CapitaOne Admissions
Northamptonshire County Council	https://www.wha	CapitaOne Education
Northamptonshire County Council	https://www.wha	CareDirector
Northamptonshire County Council	https://www.wha	CareFirst Adults (FACERas)
Northamptonshire County Council	https://www.wha	Caspar case management system
Northamptonshire County Council	https://www.wha	Fischer Family Trust Aspire
Northamptonshire County Council	https://www.wha	Free school meals / early years eligibility checker (government hosted application)
Northamptonshire County Council	https://www.wha	Outcomes Star
Northamptonshire County Council	https://www.wha	YOGRS [predicts reoffending]
Nottinghamshire County Council	https://www.wha	"hospital admissions for unintentional injuries" [Public Health Analysis Team project]
Nottinghamshire County Council	https://www.wha	"modelling to understand patient flows for various service reconfigurations regarding obesity pathways" [Public Health Analysis Team project]
Nottinghamshire County Council	https://www.wha	"population segmentation with MOSAIC data to understand spatial distribution of risk of alcohol harm" [Public Health Analysis Team project]
Nottinghamshire County Council	https://www.wha	"spatial analysis combining populations and locations of maternity services" [Public Health Analysis Team project]
Nottinghamshire County Council	https://www.wha	"use of routinely available indicator measures from PHE fingertips for monitoring and identifying work areas to focus on" [Public Health Analysis Team project]
Nottinghamshire County Council	https://www.wha	Microsoft Office
Nottinghamshire County Council	https://www.wha	Microsoft SQL Server
Nottinghamshire County Council	https://www.wha	Mosaic (Experian)
Nottinghamshire County Council	https://www.wha	QGIS
Nottinghamshire County Council	https://www.wha	Rstudio
Nottinghamshire County Council	https://www.wha	Scenario Generator
Orkney Islands Council	https://www.wha	Hazman online risk register
Perth and Kinross Council	https://www.wha	"We can advise that although the Council uses systems that will act in a manner consistent with the terms detailed as above we have not recorded information in relation to those systems using the same terms as requested."
Poole Borough Council	https://www.wha	"we envisage services will use reporting tools which form part of their caseload software packages"
Poole Borough Council	https://www.wha	Google Analytics
Purbeck District Council	https://www.wha	Capita Revenues and Benefits software
Purbeck District Council	https://www.wha	Risk Based Verification (Xantura)
Rochford District Council	https://www.wha	"Some systems will use calculations to establish eligibility, benefits for example, based upon the information manually input/provided by the customer and for specific legitimate business purpose."
Rotherham Metropolitan Borough Council	https://www.wha	Rentsense (Mobysoft)
Royal Borough of Windsor and Maidenhead	https://www.wha	Risk Based Verification
Sandwell Metropolitan Borough Council	https://www.wha	Healthy Sandwell Firmstep system
Sandwell Metropolitan Borough Council	https://www.wha	NHS Health Check Service
South Bucks District Council	https://www.wha	Risk Based Verification
South Cambridgeshire District Council	https://www.wha	"There are some examples of algorithm and/or automation including: form validation and data entry"
St Helens Metropolitan Borough Council	https://www.wha	Cognos
St Helens Metropolitan Borough Council	https://www.wha	Crystal Reports
St Helens Metropolitan Borough Council	https://www.wha	Power BI (Microsoft)
St Helens Metropolitan Borough Council	https://www.wha	PSI [children's social care reporting dashboard]
St Helens Metropolitan Borough Council	https://www.wha	SSRS
Suffolk County Council	[Not via WDTK]	"Sequel Server Reporting Solutions" [presumably SQL]
Suffolk County Council	[Not via WDTK]	BI tools
Suffolk County Council	[Not via WDTK]	Care First 6
Suffolk County Council	[Not via WDTK]	Connect Measure
Sunderland City Council	https://www.wha	Risk Based Verification (CallCredit)
Swale Borough Council	https://www.wha	Google Analytics
Swale Borough Council	https://www.wha	Google Tag Manager
Swale Borough Council	https://www.wha	Siteimprove
Teignbridge District Council	https://www.wha	Strategic Asset Management (SAM) database (ELF Software)
Torbay Borough Council	https://www.wha	"software or hardware tools to assist in dealing with large and integrated datasets"
Torbay Borough Council	https://www.wha	Business intelligence tool/s
Torbay Borough Council	https://www.wha	GIS tool/s
Tunbridge Wells Borough Council	https://www.wha	"I can confirm that our Digital Services Team use data analytics forassessing performance of the website and other online services, such as the use of google analytics for assessing website performance."
Tunbridge Wells Borough Council	https://www.wha	Google Analytics

Local Authority	FOI Link	System Name
Waltham Forest Borough Council	https://www.wha	"A 'proof of concept' project is currentlyunderway to assess the feasibility and benefits of using the Amazon WebServices (AWS) cloud platform to manage some of the Council's data."
Waltham Forest Borough Council	https://www.wha	Athena (Amazon Web Services)
Waltham Forest Borough Council	https://www.wha	Cloud BI [possibly Amazon Web Services]
Waltham Forest Borough Council	https://www.wha	Comprehend (Amazon Web Services)
Waltham Forest Borough Council	https://www.wha	DMS (Amazon Web Services)
Waltham Forest Borough Council	https://www.wha	Looker
Waltham Forest Borough Council	https://www.wha	S3 (Amazon Web Services)
Waltham Forest Borough Council	https://www.wha	Sagemaker (Amazon Web Services)
West Lothian Council	https://www.wha	Risk Based Verification (CallCredit)
West Sussex County Council	https://www.wha	Mosaic (Servlec)
West Sussex County Council	https://www.wha	Power BI (Microsoft)
Westmintser City Council	https://www.wha	Troubled Families [active programmes across England, with this authority believing it fit the definitions of our request]
Weymouth and Portland Borough Council	https://www.wha	Capita Revenues and Benefits software
Weymouth and Portland Borough Council	https://www.wha	Risk Based Verification (Xantura)
Wigan Metropolitan Borough Council	https://www.wha	"risk stratification model" [combined dataset with analysis performed upon it, doing stuff relating to school readiness, NEET, complex dependency, health (reminds me of Kent CC's Kent Integrated Dataset)]
Wigan Metropolitan Borough Council	https://www.wha	GM Connect
Worcester City Council	https://www.wha	Dennis Connect [vehicle telemetrics]
Worcester City Council	https://www.wha	Excel (Microsoft)
Worcester City Council	https://www.wha	Quartix vehicle tracking
Worcester City Council	https://www.wha	Risk Based Verification (Capita)
Worcester City Council	https://www.wha	Sunray System
Worcester City Council	https://www.wha	Worcestershire Internal Audit Shared Service (WIASS)
Wycombe District Council	https://www.wha	National Fraud Initiative (NFI)