PARENTAL SOCIAL LICENCE FOR DATA LINKAGE FOR SERVICE INTERVENTION:
A UKRI-FUNDED RESEARCH PROJECT
https://generic.wordpress.soton.ac.uk/parentdata/

Working Paper:

Freedom of Information Requests on the Use of Data Analytics in Children's Services: Generating Transparency

Val Gillies and Bea Gardner
with Rosalind Edwards and Sarah Gorin

Openness, accountability and transparency has emerged as a theme in the UK government's championing of the use of data linkage and algorithmic analyses. For the most part, this theme is centred on openness between services in sharing administrative records and accountability as legal compliance. There is, though, some equation of transparency with public trust, with recommendations that people should be told about what is happening to their data and how it is used. A Local Data Accelerator Fund for Children and Families has been initiated, where local authorities can bid for funding for data sharing and matching projects that support identification of families for 'earlier intervention before risk escalates'. But there is no central record available of which local authorities are doing what when it comes data sharing, operational linking and matching, and applying predictive analytics to families' administrative data. In an effort to map the state of play among British local authorities we submitted Freedom of Information requests to local authorities.

We submitted a Freedom of Information Request for documents relating to previous or current uses of data analytics, predictive analytics, or algorithmic automated systems used for processing, risk assessment, scoring systems or automated decision making in relation to child and family services within the jurisdiction of the local authority concerned. We asked for information about:

1. any use of data analytics and or artificial intelligence systems to support children's service provision:
2. any previous use of data analytics and or artificial intelligence systems to support children's service provision;
3. any intended use of data analytics and or artificial intelligence systems to support children's service provision;
4. any briefing notes, reports, or summaries relating to the use or potential use of data analytics and or artificial intelligence to support children's service provision;
5. any meeting minutes, reports or presentation materials on the efficacy or impact of data analytics and or artificial intelligence systems in relation to children and their families; and
6. any third-party consultants or suppliers of data analytics or artificial intelligence including any contracts, memoranda of understanding, or other agreements (which may include but not exclusively data sharing agreements).

In themselves, our Freedom of Information (FoI) requests yielded little in the way of concrete information on the use of data analytics in children's services. Nonetheless, it has been

a revealing process, not least in uncovering the lack of transparency councils maintain when implementing or developing digital technology. Aside from anything else, we have a demonstration of how ineffectual the FoI system is. We have limited details, but the constrained information we did glean led to a trail of key players, quangos and organisations driving data led agendas. Our report here brings the FoI information (or lack thereof) that we received into engagement with relevant information that we have been able to source from websites and accessible online materials. In other words, we have had to link together information to enable us to generate at least some transparency where little exists.

Key issues summary:

- Local authorities may avoid or obscure transparency about their use of data analytic systems with predictive capacities in response to FOI enquiries, but other information shows that their use is common. While predictive analytics is not a common tool for triaging child protection referrals, many English councils use some form of family tracker.

- Local authorities utilise a variety of digital tools to classify, flag, and target families and young people, but can claim FoI exemptions to avoid transparency about their use of data warehousing and data lakes, and the involvement of private data analytic companies.

- There is a shift in policy focus and data practices, away from early years intervention towards criminal justice and a concern with identifying poor families as criminogenic. There is little evidence of transparency for families about use of their data for classification and intervention, or that data analytic practices are GDPR compliant.

Ambiguous terms

Although we recognised that language would be a challenge in formulating an effective FoI request, we did not anticipate the extent to which it would be used to avoid transparent provision of information. We asked for details of any use of data analytics, predictive analytics and algorithmic automated systems used for processing, risk assessment, scoring systems or automated decision making in relation to child and family services. All English councils participating in the Troubled Families Programme (now called the Supporting Families Programme) are by necessity using data analytics in that they must collect and link together data sets to identify 'troubled' households and claim payment-by-results from central government. Yet only 76 out of 149 English councils responded that they used data analytics. The remainder claimed that their systems did not meet our definition or responded with a straight 'no' to all of our questions about their use. Councils defining themselves outside of our criteria could adopt semantic ambiguities. For example, Chester West stated: 'We are defining 'data analytics' for the purposes of this FoI as automated data processing with algorithms', before answering 'no' to all the questions. Others responding 'no include councils who are using Low Income Family Trackers (see below).

Online searching revealed that those answering 'no' in response to our requests for information about any engagement with data analytics either defined their family work outside of children's services or used the term 'data matching' instead. Data matching can range from simple merging of Excel spreadsheets to 'data lakes' (pooling a range of data formats – see below) and the production of dashboards. Information on the Local Government Association and the UK Government Digital Transformation Fund websites, and the sites of companies like Xantura, make it evident that many of these councils in fact use systems with predictive analytic capacity. For example, Achieving for Children (a social enterprise providing services for children in several local authorities) responded to our FoI that they used 'some basic data monitoring' but that 'we do not currently use nor have we previously used any data analytics, predictive analytics and or artificial intelligence systems to assist with this work'. Yet they have used business intelligence technologies on a range of projects using predictive analytics/algorithms, as noted on the UK Authority ('Digital data and technology for the public good') website: https://www.ukauthority.com/articles/the-digital-boost-at-achieving-for-children/.

A majority of the FoI responses deny use of any automated decision making. Thurrock, for example, emphasised this in its response, stating:

> We do not use any data for predictive analytics or AI in the provision of Children's Services. We currently use data only in relation to verifying families that meet the criteria to enter into the Government's Supporting Families programme (previously the Troubled Families Programme). NO AUTOMATED DECISIONS are ever made during any of this process. (original emphasis)

Nonetheless, Thurrock features as a case study on Xantura's website in their adoption of the OneView service:

> [OneView] enabled all of the data extraction to be automated and matched using a unique pseudonymisation approach and embedded information governance. A range of operational tools were then delivered enabling Thurrock to see all their data at an aggregate level to identify relevant cohorts & risks as well as identifiable case summaries

for front-line staff and payment dashboards for management. https://xantura.com/thurrock-council-supporting-families/

Under the current General Data Protection Regulation (GDPR) there is a transparency requirement to explain fully the use of automatic systems to anyone affected and to gain explicit consent to use of the data:

> The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her – Data Protection Act, Section 4 (4).

Councils, then, appear to be relying on a contingent interpretation of this stipulation. While 'automatic' would ordinarily be used to describe as an activity conducted independent from human engagement, councils appear to frame it as a process free from human involvement, an improbable scenario short of a technical singularity. Profiling, defined in the GDPR as 'automated processing of personal data to analyse or predict' is not recognised as subject to the requirements of the Act even though it is clearly intended to be. North Yorkshire County Council appeared mindful of these sensitivities in their response to our FoI:

> There are no systems, augmentations or AI implementations that replace human decision making processes. We have developed some views of data and activity that compliment a human process by highlighting certain attributes to decision makers.

Councils also side-stepped the term algorithm and the concept of AI. Almost all councils engaged in predictive analytics denied they were using algorithms, yet it is hard to envisage predictive models that are not built from algorithms. Openness, accountability and transparency become circumvented through these and other practices.

Evasive practices

Transparency was also constrained through a lack of provision of information requested in our FoI. Hardly any councils acknowledging use of data analytics and predictive algorithms provided any papers or documents relating to their use. Looking across all the responses we received, regional differences between England and the rest of the UK are evident. Few councils in Scotland or Northern Ireland appear to be using what we would recognise as data analytics (slightly more in Wales), yet many sent documents relating to databases and individual risk scoring tools in an effort to meet our request. Local authorities in England however, are far more likely to have adopted data analytics (given the Troubled Families Programme was instituted in England only), but not to disclose this usage.

A common practice was to send insufficient or incomplete information. A number of English councils answered our question as to whether they were using data analytics, predictive analytics and or artificial intelligence systems with just one 'yes'. This meant that we had to send a follow up request to establish which of these they were referring to (the answer was usually 'data analytics' with no further clarification). Some councils sent random documents with little relevance to our questions. For example, Islington stated that they used data analytics but met our request for documents by sending a family intervention scoring questionnaire and a link to their Ofsted report. They neglected to mention their construction of a 'data warehouse' (storing structured data – see below) and significance of this to their work with families.

Hammersmith and Fulham sent an isolated appendix from a report that related to predictive modelling which was incomprehensible on its own. They gave no indication of the title

making it impossible to request the report itself. They also sent a page from a previous FoI inquiry about predictive modelling in children's services (minus the original question). We managed to locate the original request, which we found had gone to an internal review because so little information was disclosed. The complaint was partially upheld but the requester gained only a data protection review document and the front page of an evaluation report. It is difficult to imagine that Councils thought that they were providing transparency through such replies to our FoI.

Councils could declare exemptions under the FoI Act to avoid supplying the information we requested. A section 12 exemption on the grounds that the cost of compliance exceeded the permitted threshold was commonly asserted. For example, Stockport claimed that:

> We currently have 154 Tableau Workbooks covering Early Help and Children's Social Care, which consist of 624 dashboard views. To confirm which systems are used to support this provision and to extract any such information would require the relevant Council service area(s) to manually search through all 624 dashboard views.

Section 43 Commercial Confidentiality clauses were also used. Ealing disclosed that its predictive model was built in commercial agreement with Hammersmith and Fulham and argued that providing the requested information could prejudice their pending or future negotiations. Similarly, Lincolnshire did not feel it appropriate to share what it judged to be sensitive information about a 'joint agency piece of work'. Oldham claimed a section 43 exemption from even disclosing the name of company that was setting up their data warehouse and predictive analytics capacity, while Stoke on Trent would not disclose the name of the company rolling out the Power Business Intelligence data analytics tool for them. Islington and Hackney both claimed commercial confidentiality to avoid releasing any evaluation reports relating to their predictive models. Involvement of commercial companies seems counter to transparency about data linkage and predictive analytics.

In some cases, we submitted revised and follow-up requests in attempts to tease out information from the local authority. For example, Southwark rejected our question about any future use, stating:

> Please note that this question, in its current form, does not constitute a Freedom of Information request and is therefore not answerable as part of this request. The Freedom of Information Act 2000 provides an applicant with a right to access specified information 'held' by a public authority at the time of the request (Section 1(1)). The Act does not, however, require the Authority to create information in response to a request received or provide any statements of intent with regard to its day-to-day operations.

We submit a restructured request asking about future *plans* but the response was slightly ambiguously worded: 'We do not hold any such information'.

After a long delay and a request for us to clarify what 'spectrum' of analytics we are referring to in our FoI, Tower Hamlets replied with an uninformative 'yes' and citing two exemptions to avoid transparency. In addition to the cost exemption, they claimed a Section 21 exemption on the grounds that the details of the suppliers were already available in the Council's Contact Register. The register they referred to lists hundreds of council contractors, providing only names and worth of the contract. Our efforts to clarify at least what type of data practices Tower Hamlets are using took months. After a prompt it was established the request had been

lost. It was then found again. Eventually the council responded that they used both data and predictive analytics, but no further details were forthcoming.

The data 'solutions'

Where some transparency at least was provided, our FoI request revealed the wide variety of tools and companies employed by local authorities. Microsoft Business Intelligence (MS BI) was the most common toolset cited, although we gained little detail as to how this is being used across different authorities and for what purposes. In its most basic form MS BI can be used to provide a broad visualization of service use and cost, and as a form of performance management. Depending on the version (or functional upgrades) purchased however, it can be used for data warehousing (or cloud lake functionality) and predictive analytics. At this point it is useful to note the difference between 'data warehouses' and 'data lakes': a warehouse tends to store structured data (SQL data in files and folders) whereas a lake pools structured, semi-structured and unstructured data. A good example of how MS BI is being used comes from Dorset County Council who have written a case study for the Local Government Association of how this tool can 'enable local professionals to identify potential difficulties for individual children before they become serious problems: https://www.local.gov.uk/sites/default/files/documents/Dorset%20Case%20Study%20Digital%20Transformation.pdf. These MS BI dashboards, providing sensitive family details and child level flags, are accessible to 279 school staff across the Dorset County Council and a further 132 Family Workers (not including IT staff and managers). According to the case study the dashboard is most often used by schools. Our FoI established that Dorset Children's services are looking to make greater use of AI in the future.

Tableau is another common tool. In contrast to MS BI it does not warehouse or pool together data but rather connects to any number of databases and queries them directly. Quite how it is being used is opaque, but in the following online example: https://www.tableau.com/solutions/customer/pennine-nhs-foundation-trust-holistic-childrens-care, Tableau is connecting children's services databases to the NHS Trust's warehouse. Tableau also has the capacity to undertake predictive analytics using machine learning, albeit there is no evidence that it is necessarily being used in the way.

Qlik is another commonly used software tool, often in collaboration with MS BI. Although councils provided us with few specific details, it appears to be used by Bristol (their Insight hub uses Qlik as well as MS BI), Hackney, and Camden. Qlik supports profiling, risk scoring, fraud detection and predictive policing. Councils are also using some version of Liquidlogic, a software system designed for local authority social care and education services. This offers a 'one view' case management dashboard, though it appears that Surrey experienced difficulties trying to customise it: https://www.communitycare.co.uk/2019/03/07/chaotic-implementation-liquidlogic-case-management-system-surrey/. A variety of other tools that may be used in conjunction with these systems include cloud systems like Azure, Cloudwick or Amazon Web Service, SAP, to help integrate different data sources; other add-ons support text mining/scraping (e.g. Rapidmine, Provalis) and predictive analytics/machine learning (e.g. KNIME, Alteryx).

While some councils have invested in data scientists to develop, implement and/or manage these tools, others have bought in private companies. Councils with in-house expertise appear most likely to be implementing data lakes, predictive modelling and the Internet of Things (e.g. Bristol, Brent, Camden, Essex, Kent, Hammersmith and Fulham). In some cases these

operate as separate companies within the given council and can contract to provide the models to other councils (e.g. Ealing is purchasing a predictive model from Hammersmith and Fulham).

As noted, councils could claim section 12 exemption to avoid transparency on the involvement of private companies. Of the few councils that provided names, Xantura was mentioned most often and supplies councils with a wide range of surveillance and predictive modelling projects. The only details we got on their involvement from the FoI however, came from Barking and Dagenham, who sent a commissioning document, and Southwark, who sent a decommissioning document detailing the poor accuracy Xantura's system had delivered. The Barking and Dagenham document was revealing in highlighting the financial support that council received. The contract with Xantura was worth £1.25M, but this was subsidised by London Ventures as part of Capital Ambition, an organisation set up by central government to promote innovation in London local governance. London Ventures is run as a partnership with Ernst Young and in turn has partnered with Xantura (meaning there was no tendering). Xantura are often partnered with Ernst and Young (a multinational professional services network) and are sometimes referred to as EY Xantura. There is in fact a proliferation of funding streams and quangos supporting councils to adopt data analytics. The Innovation Fund, the London Office of Technology and Innovation (LOTI), the Digital Information Fund, the London Challenge Programme, Social Finance, and the Local Digital Accelerator Fund are all enabling councils to adopt and develop data analytics in children's services. It is not clear how much any subsidies are worth though. Companies can also incentivise the use of their products through promising to find part of their own fee. For example, the commissioning document for Barking and Dagenham states:

> There currently exists a budgetary provision of £550k within Community Solutions (Transformation) Capital IT funding to cover most of the first two years of the EY Xantura costs. The balance is to be funded from the savings identified by EY Xantura themselves.

H&F, the company spun out of Hammersmith and Fulham similarly promise not to take a fee unless it can identify a £50k saving.

Other private companies briefly mentioned in the FoI responses or elsewhere include: Sentinal, Agilisys, Predict X, IBM, IMPOWER, Mastodon C, and Thoughtworks. Devon and Cornwall used small local firms (Peak Indicators and Headforwards). While there are a large variety of tools and companies currently in use, the direction of travel seems to be towards an amalgamation of datasets across councils and beyond. As we noted earlier, the £7.9 million Local Data Accelerator Fund from central government is designed to 'improve joined-up working between councils, NHS trusts and police', while other funding sources require councils to apply together on their data systems.

Protection and prevention

Our analysis of the FoI responses combined with the information we have sourced from elsewhere points to a clear shift in service priorities away from early intervention for parental education towards child protection and crime prevention. The prior stress on linking parenting skills to social mobility is now muted, with rationales for data innovation focusing almost exclusively on the pre-emption of problems rather than on the maximisation of children's future potential. References to 'early intervention' remain but are used to emphasise cost savings and harm prevention. The introduction to the Digital Transformation case study produced by Dorset County Council illustrates this change in meaning:

> Evidence shows that prevention and early intervention give better results for children and young people because issues are addressed before they become serious. Hence, over the long term, early intervention is much less costly than late intervention.

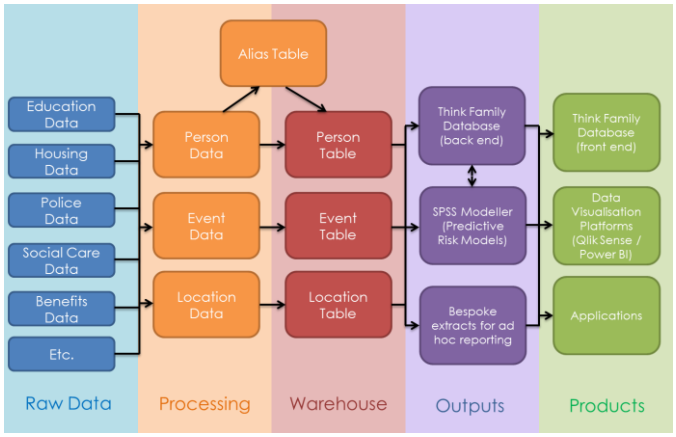
Dorset's document sets out an ACE-style (Adverse Childhood Experiences) pen portrait of 'Tom', a boy who witnesses domestic violence, struggles at school, and then grows up to beat his girlfriend and child, ending up in jail:

> In response to this Dorset asks "What would have happened if…" they could use business intelligence to predict when early intervention could make a real difference to such scenarios?

The question is answered through reference to a dashboard project designed to identify problems 'early', so they are 'nipped in the bud'.

Despite the emphasis these projects place on acting early, the age of children targeted for intervention has moved significantly upwards and away from the previous policy preoccupation with under-fives. The Dorset Dashboard project monitors children between the ages of 5 and 15, while predictive analytics projects tend to profile over 10s. This likely reflects the origin of these data projects in the Troubled Families Programme, but also the preoccupation with crime reduction and antisocial behaviour that was used to justify its introduction. Most active predictive modelling projects are designed to prevent gang crime, county lines, sex trafficking and antisocial behaviour, domestic violence, homelessness or fraud. A typical example is Bristol County Council's 'Insight Bristol' project, an interagency data warehouse and analytics hub run by Avon and Somerset Police (known for their pioneering use of predictive policing).

Insight Bristol are operating four key risk models targeting 'children' at risk of 'criminal exploitation', 'sexual exploitation', 'becoming NEET' (not in education, employment or training) and 'violence' (perpetrating rather than falling victim to it). Bristol provided us with documents in response to the FoI but most of them proved to be old and none listed exactly what data was being used. The local authority did provide the broad diagram below but this is likely to be out of date given their current website mentions collecting data on 'families where children need help' and 'parents and children with health needs'.



A lack of transparency is a similar story in other councils. According to a document supplied by Brent, their predictive analytic journey began with a commitment to reduce child sexual exploitation. We are told that a report by the Children's Commissioner prompted an

urgent question as to whether there was a child sexual exploitation (CSE) problem in Brent. Datasets had to be merged to create a 'vulnerability index' so the question could be answered. This narrative is illustrated with insensitive tabloid headlines about Asian gangs trafficking girls and concludes with the discovery that the most at risk children are 'not just involved in CSE but Gangs, County Lines, Drug Running'. Brent found that children tended to progress up a pyramid of risk, leading the council to ask 'If we have the means of identifying young people early on and *engaging with them to prevent harm*, should we not do so?' (their emphasis). After 'processing 220 million data records' they discovered that children known to social care, excluded from school, reliant on benefits and living in gang active areas are most at risk. Brent now has a dashboard that identifies and tracks these young people.

Brent have also built on their single view dashboard by developing a data lake, with the aim of 'advancing analytics and machine learning and applying insights to various other issues, including preventing rent arrears and better identification of unlicensed properties' (Brent Data Led Service Delivery). The lake will include structured, semi structured and unstructured data (emails, documents, PDFs) as well as binary data (images, videos, audio) and will, according to Brent, support connections with the Internet of Things (everyday devices, machines and objects with internet connectivity):

> Applied to social care, it allows us to monitor ourselves and our environment in real-time – for example, heart rate, temperature, blood sugar, the list is endless – enabling professionals to decide if any next steps or intervention is needed.

Some council materials on digital innovation make for a lack of transparency even while purporting to provide it. For example, Essex County Council (in collaboration with Essex Policy and the University of Essex) has a Centre for Data Analytics (ECDA) with a public website that is the core tool of their 'open data' approach. The website conveys data analytics as beneficial and progressive (including an interview with a data analyst titled 'Everything is Awesome' and a quiz to find out what kind of analyst you are). Yet it is difficult to access much meaningful information on data analytic practices and service intervention. One pilot project run in collaboration with the data analytic company Predict X sought to identify children at risk of not being school ready as an example of how predictive analytics can ensure children get the best start in life. Yet there is no indication that any extra support was directed towards the identified families. Two other pilot projects on domestic violence and gang crime/county lines resulted in more concrete actions, including the establishment of a violence and vulnerability unit to process and action the data insights on 16000 risky households.

Hammersmith and Fulham's data analytics spin off (H&F Business Intelligence) operates a similarly focused predictive model, declaring on their website that the identification of young people at risk of offending is 'one of our obsessions'. Initially they were using predictive analytics to identify children at risk of entering local authority care, but state this is no longer operational in their local authority. Rather, the focus now seems to be criminogenic. For example, a council collective made up of Hammersmith and Fulham, Westminster, and Kensington and Chelsea participated in a data led project called 'On Track', aiming 'to reduce the number of children aged 10 years and over entering the care system and *being at risk of custody'* (our emphasis).

As part of the shift towards criminal justice, there is a concern with identifying poor families as potentially criminogenic.

Predicting poverty

The focus of our FoI on children's services resulted in us missing one of the most extensively used forms of predictive modelling applied to families. Big Brother Watch has published a detailed analysis of the use of data analytics by councils and points to widespread use of Low Income Family Trackers (LIFTs). Developed by the company Policy in Practice, these trackers are used to identify disadvantaged households in order to target them for interventions (e.g. advice on debt, invitation to a weight loss clinic). The trackers process large amounts of linked data to categorise families as: in crisis, struggling, at risk or coping.

Xantura's OneView product can be used in a similar way depending on the package invested in. As well as the children's social care package, there are versions that predictively model risk of debt and financial vulnerability, of homelessness and risks associated with Covid. The programme can handle structured and unstructured data, and can scrape from the web in an effort to identify domestic violence, mental health conditions, medical conditions and disability. Big Brother Watch found that in Shropshire, Covid OneView scraped data relating to individuals' anger issues, dangerous dogs and history of unsafe sex (https://www.pressreader.com/uk/daily-mail/20201128/281754156880560).

Risk based verification is another widely used form of profiling of poor families in particular. A range of characteristics are scanned to rank benefit claimants as at high, medium or low risk of fraud (those flagged as high risk endure extra checks and telephone calls). Xantura are again key players, yet Big Brother Watch found them to be amongst the most secretive of data analytics companies. It is difficult to see how public trust can be built on transparency about data linkage and preventive analytics under these circumstances.

Final reflections

The lack of transparency about the use of data linkage and predictive analytics in children's social care makes it difficult to map a firm picture of local authority practices in the field. It does appear that few councils are using specially designed predictive modelling to inform child protection work, but it is also clear that a wide a range of data 'solutions' are drawn on to classify, flag, target and intervene in disadvantaged families and their middle age or teenage children.

The success of our FoI attempt for transparency was restricted by the shift that local authorities have made away from early years services over the last decade. Children are no longer accorded the status of 'index' case in local authority databases and traditional children's services seem largely to be giving way to predictive policing. We might have been more successful it we had asked specifically about data on families, given that it looks as though data collection and categorisation of households has become routine.

Rather than transparency, there is little indication that families are being made aware of how they are being categorised, and if they are targeted they may not realise why. In some cases they may not even realize they have been targeted at all (e.g. flagged for a fraud investigation or sent what might look like a generic information leaflet). It is hard to see how all this 'innovation' can be interpreted (however creatively) as GDPR compliant currently, or within the framework of the Regulation of Investigatory Powers Act. Profiling is used by councils without the knowledge or consent of families. Further, data relating to children can be retained for a long period of time (e.g. Bristol keeps records relating to school exclusion, truancy and SEN (Special Educational Needs) status for 25 years, EHCP (Education and Health Care Plan) applications for even longer and child protection records for 75 year plus). While there is a case for archiving this data, there is no transparency about whether or not it is being stored and mined as part of council

warehouses or lakes. If so, this would contravene the current GDPR's requirement that data should be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes.

The UK Government recently has proposed reforms to current data protection legislation, attempting to make legal and cultural limitations on data sharing, linkage and analytics subject to more flexibility, and to expand data processing and sharing provisions to private companies. The consultation document refers to a need for high levels of public trust and transparency in the data regime. It is hard to see how this will be achieved when transparency is starting from the opaque, ambiguous and even evasive base that our FoI request endeavours revealed.