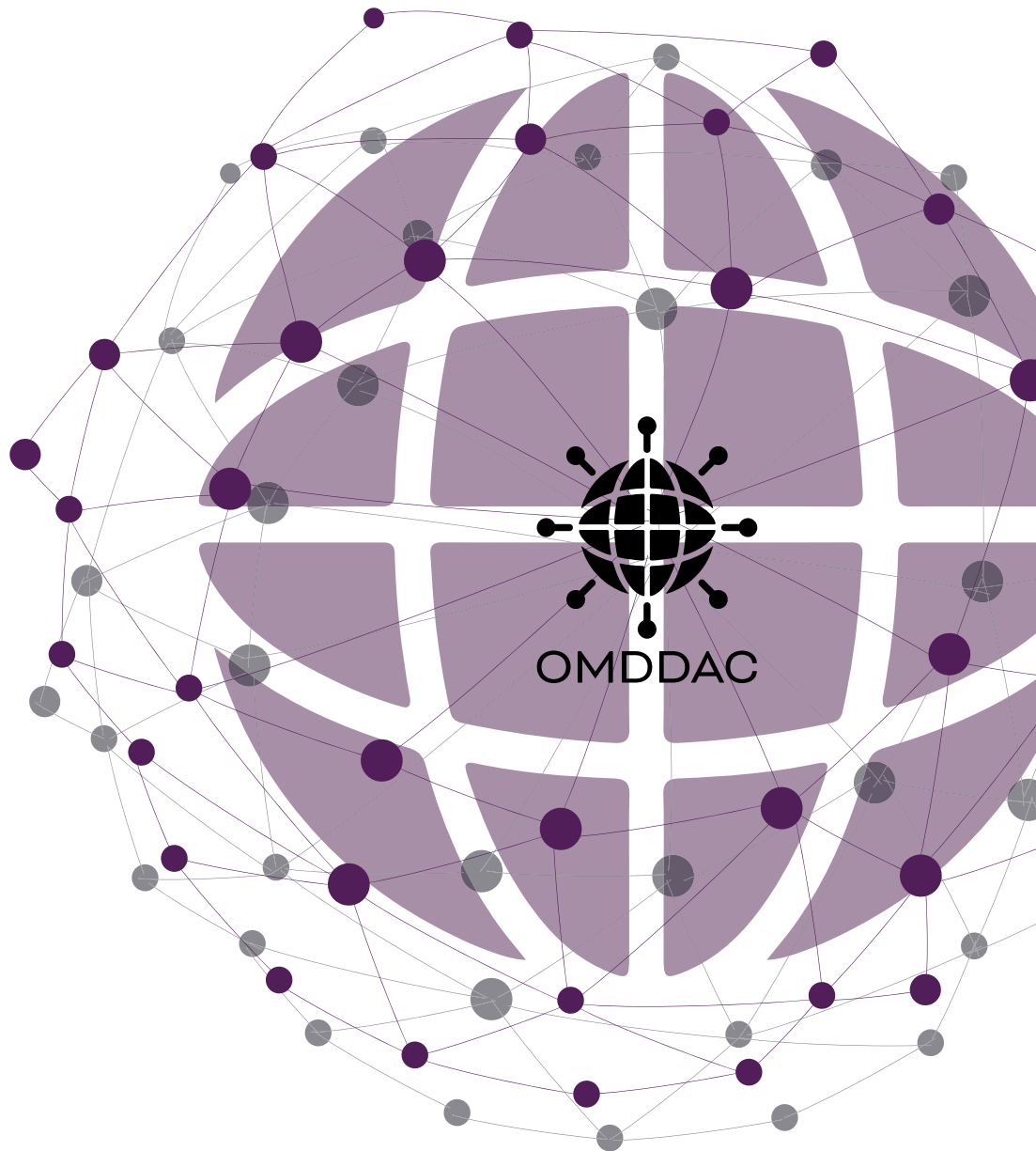


Data-Driven Responses to COVID-19

Lessons Learned

OMDDAC Research Compendium



Rachel Allsopp, Claire Bessant, Sneha Dawda, Keith Ditcham, Charlotte Emmett,
Matthew Higgs, Ardi Janjeva, Guangquan Li, Selina Sutton, Mark Warner

Principal Investigator: Marion Oswald

October 2021

Contents

4	Executive Summary
7	Introduction
9	1. Data Foundations
9	1.1 Data Quality
12	1.2 Data Infrastructure
13	2. Data Skills
13	2.1 Data Literacy
16	3. Data Availability
16	3.1 Data Sharing During the Pandemic
17	3.2 Public Concerns
27	4. Responsibility
27	4.1 Transparency
31	4.2 Public Engagement and Trust
34	Concluding Recommendation and Next Steps
37	Appendix I - Methodology



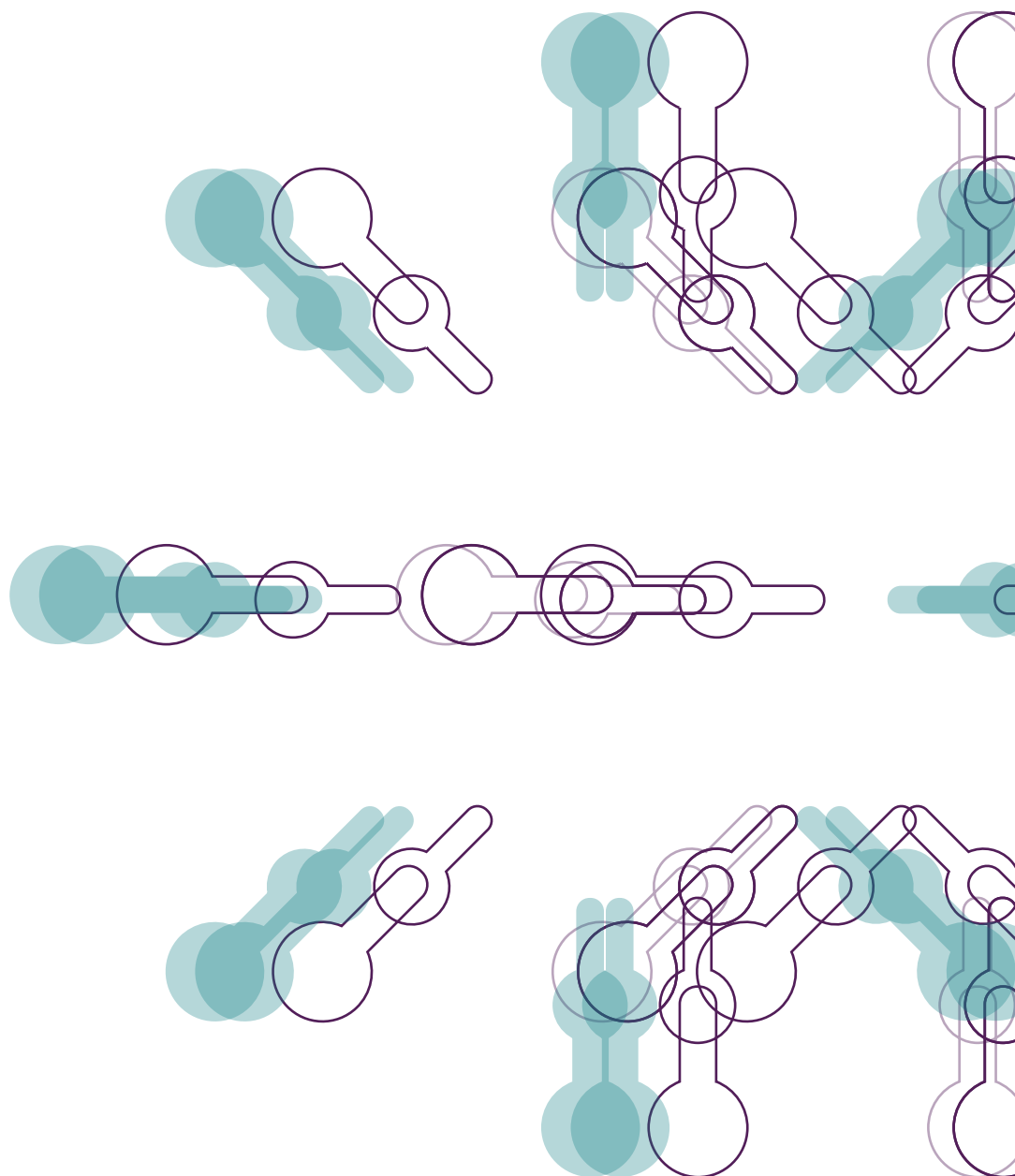
Acknowledgements

The authors would like to thank all those who took part in our interviews, surveys and the agenda day as part of this project, without whom the research would not have been possible. The authors would also like to extend thanks to Investing in Children for organising and facilitating the agenda day and accompanying questionnaire.

The authors are very grateful to Sacha Babuta and John Rooksby for kindly reviewing earlier versions of this report and providing very helpful feedback. The authors would also like to thank Roots and Wings for their expert design services in producing the report.

The authors would like to thank Hugh Oberlander for his invaluable support as project officer throughout the OMDDAC project.

Finally, the authors would like to extend thanks for the support of our project partners Lord Jonathan Evans (Baron Evans of Weardale), the Ada Lovelace Institute, the Centre for Data Ethics and Innovation and medConfidential, and to the Arts and Humanities Research Council for funding the OMDDAC project (grant reference number: AH/V012789/1).



Executive Summary

Funded by the Arts and Humanities Research Council under the UKRI COVID-19 Rapid Response call, the Observatory for Monitoring Data-Driven Approaches to COVID-19 (OMDDAC) is a collaboration between Northumbria University and the Royal United Services Institute (RUSI). This project has involved a multi-disciplinary team of researchers (with expertise in the law on technology, data protection, and medicine as well as practical ethics, computer science, data science, applied statistics in health, technology and security studies and behavioural science) to investigate the legal, ethical, policy and operational challenges encountered in relation to key data-driven responses to the pandemic.

The COVID-19 pandemic has accelerated the consideration of several priorities in the data and technology space, which are reflected in the UK Government's present strategies. The National Data Strategy, in particular, pledges to take account of the lessons learned from the COVID-19 response and draw upon the UK's values of **transparency, accountability** and **inclusion**.¹

Seeking to inform the lessons learned from the pandemic, the project used a mixed-methods research design that included case study analysis, interviews with key stakeholders (individuals with relevant expertise and/or experience in relation to the data-driven pandemic response), representative public surveys, and engagement with young people through a children's rights charity. OMDDAC has published four snapshot reports focused on data-driven public policy, tech-driven approaches to public health, policing and public safety and key findings from the public perceptions survey. The emerging issues identified in those reports align closely with the four pillars of the National Data Strategy, which form the framework for this final project report:

1. **Data Foundations** (data quality issues and infrastructure);
2. **Data Skills** (data literacy of decision-makers);
3. **Data Availability** (data sharing); and
4. **Responsibility** (law, ethics, transparency, and public trust).

Data Foundations: in our interviews, stakeholders consistently highlighted significant challenges relating to data quality and operational capability. We found that a *"thirst for information"* across policing and government during the pandemic led to the creation of new data sets with **varying levels of data reliability**. Interviewees also talked of a wide **diversity in technical and operational capability** across the public sector. These findings echo those of the National Data Strategy which identifies obstacles to the effective sharing and use of data arising from inconsistencies in standards and data infrastructure across government. We therefore recommend:

- the establishment of a national public sector data standards framework to address inconsistencies and enable better interoperability of public sector data;
- that corrective action be taken to address quality issues with data across the public sector;
- the collection and incorporation of qualitative data in support of quantitative data, for informing sound decision-making;
- that where AI and automated decision-making tools are employed to feed into decisions affecting individuals, they should continue to be subject to meaningful human oversight mechanisms;² and
- that there is consultation across the public sector to determine the ICT infrastructure and resource requirement to facilitate appropriate and secure data sharing and analysis.

¹ Department for Digital, Culture, Media and Sport (2020) 'UK National Data Strategy'.

<https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy> accessed 14 September 2021.

² In the context of our research, we limit AI to the use of data analytics and machine-learning algorithms.

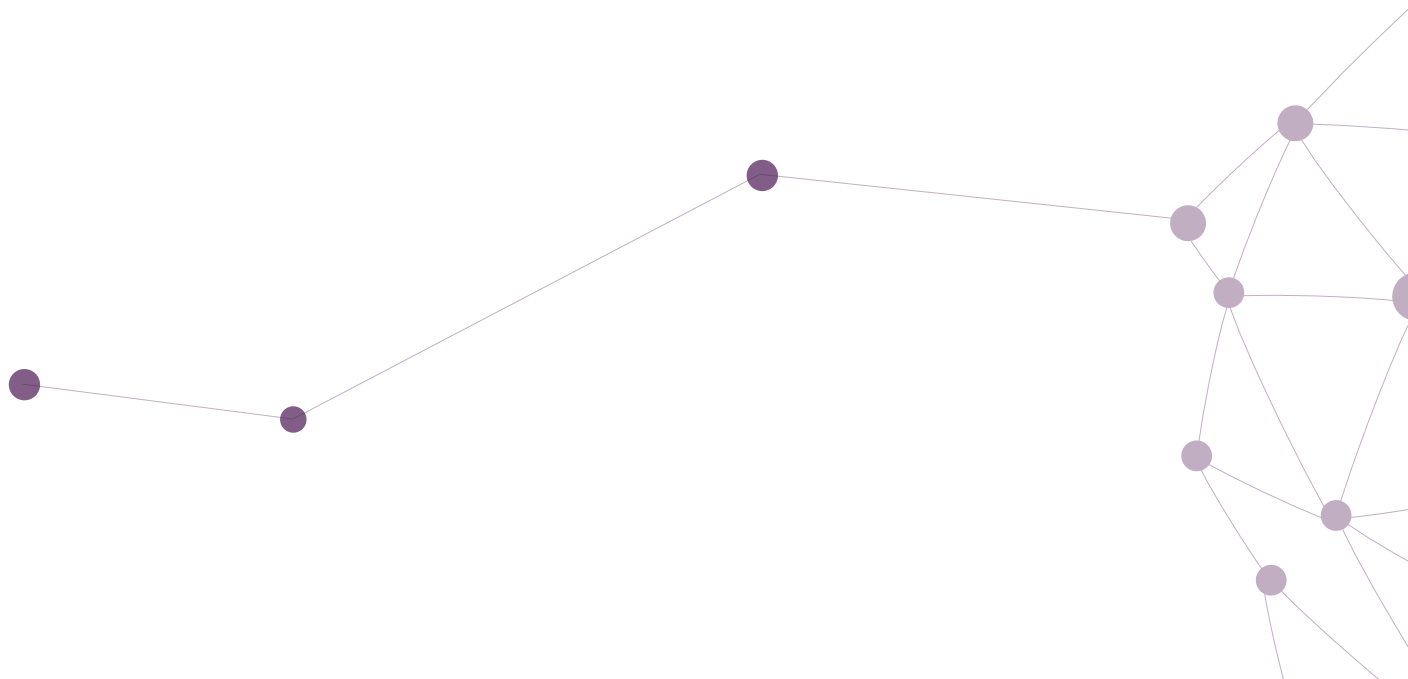


Data Skills: the pandemic exposed important skill gaps related to data-driven approaches. In particular, our findings identified a gap in the **data literacy capabilities of government leaders and decision-makers** at a central and local level. Concerns were expressed about a 'breakdown at the interface between analysing data and deciding what to do with it', calling into question whether decision-makers are **properly equipped to evaluate and utilise data-driven outputs, taking into account relevant limitations, assumptions and uncertainties**. We recommend that:

- decision-makers and policy-makers in the public sector be required to undertake data literacy training and development; and
- there is more active encouragement of interdisciplinary collaboration between specialists in the methodology (including epidemiologists, data scientists and mathematicians) and domain specialists (including behavioural scientists as well as, critically, law and governance specialists).

Data Availability: the pandemic saw unprecedented levels of medical- and non-medical data sharing as well as innovative ways of doing so. The **sharing, linkage, and repurposing of datasets from a diverse range of sources has been a central tenet of the national pandemic response**. OMDDAC's research showed that the sharing of sensitive health data with bodies like the police, for example, has evoked some public concern. It is therefore crucial that data acquisition, sharing and processing is guided by **robust information governance requirements that are transparent to the public and that any restrictions and limitations are clearly laid out in law**. It is evident that there is work to be done, especially as our research indicated that data sharing between certain organisations is likely to be of concern to the public. To **create an environment where data sharing and analysis is trusted and accepted by the public**, we recommend that:

- data acquisition, sharing and processing is subject to robust, 'end-to-end' information governance requirements that are made transparent to the public, with clear restrictions and limitations set out in law and regulation;
- the design of data-sharing initiatives aims to maximise public acceptability by:
 - being transparent about what data is being shared and with whom;
 - avoiding assumptions about whether the public is comfortable with their data being shared across all sections of the public sector;
 - avoiding a 'one-size-fits-all' approach to data sharing; and
 - providing options for how much detail is shared.



Responsibility: Our research identified two key issues which are vital for the responsible use of data: **transparency** and **public engagement**. Stakeholders highlighted a need for greater transparency firstly with regard to the fundamental capabilities of data **to ensure that the outcomes of statistical modelling are not presented, or perceived as, 'the truth'**. In addition, our research highlights a need for greater levels of transparency and public engagement with regard to the ways in which data is used. It is evident that there exist real public concerns around specific types of data sharing that need to be addressed directly and we encourage a **new public conversation post pandemic** to ensure that the public are better informed, educated and consulted regarding the use of their data. The views of young people – who have been described as ‘the hidden victims of COVID-19³ – must also be incorporated into this conversation. The **young people who participated in our research felt that their views had not been considered** enough during the pandemic and suggested several ways in which the Government could do more to inform and engage more effectively with this vital section of the population. We recommend that:

- future policy decisions informed by data modelling are accompanied by explanatory statements, which are published in a format accessible to the public and detail the relevant limitations, assumptions and uncertainties. The form and content of the explanatory statements should be decided in consultation with multiple sections of the public, including children and young people;
- the public should be informed in a transparent and accessible manner about which organisations have access to their data;
- where Memoranda of Understanding are agreed between enforcement bodies and health agencies for the sharing of sensitive health data, the terms and conditions must be made as transparent to the public as possible;
- the Government Office for Science should initiate a new public conversation to consult with, inform and educate the public regarding the use of data; and
- public consultations must be inclusive of all members of the public, including under-represented groups and children and young people, employing multiple strategies of engagement.

The findings of the OMDDAC project highlight the **central importance of data quality and integrity, robust information governance mechanisms and public transparency for creating an environment where data analysis and sharing can be trusted and accepted in an emergency context**. We therefore recommend, in conclusion, that these obligations are reinforced by a regulatory oversight mechanism whereby **an appropriate oversight body** is appointed during emergency situations to ensure that:

- use and sharing of data by central government and public bodies for operational and public policy purposes serve the public good;
- data use and sharing are supported by appropriate public transparency about data sources and methods of analysis;
- data analysis is conducted with quality and integrity; and
- data sharing is overseen by independent governance and is subject to appropriate controls, conditions and time-limitations.

³ Barnardo's (2021) 'Supporting the hidden victims of Covid-19: Lessons from the First Wave.

<https://www.barnardos.org.uk/research/supporting-hidden-victims-covid-19-lessons-first-wave> accessed 16 September 2021.



Introduction

Research Rationale

Northumbria University and the Royal United Services Institute were funded by the Arts and Humanities Research Council to conduct a research project entitled 'The Observatory for Monitoring Data-Driven Approaches to COVID-19' (OMDDAC). This report is the culmination of a year's research, comprising: i) interviews focused on the experiences of key stakeholders regarding pandemic data-driven policymaking, technology and public health, and policing and public safety; ii) original quantitative research in the form of public surveys; and iii) engagement with children and young people, being one of the under-represented voices in the public debate on data and COVID-19. The aim of the project was to establish a mechanism for monitoring the key developments in data-driven approaches to the COVID-19 pandemic, shining a light on some of the most important legal, ethical, regulatory, and policy challenges that have arisen in this period. We present the lessons learned during this period in this report to inform policymakers' responses to similar challenges in future.

The pandemic has accelerated the consideration of a number of priorities in the data and technology space, and these are being reflected in the present strategies of the UK Government. In particular, the National Data Strategy pledges to take account of the lessons learned from the COVID-19 response and draw upon the UK's values of transparency, accountability and inclusion.⁴ This is imperative as the pandemic has fundamentally elevated the perceptions of those working with data every day of what can be achieved when there is increased urgency and lower barriers to sharing and engagement. The public debate on COVID-19 and data can be expected to pivot away from identifying the key data-driven responses to the pandemic, towards identifying the outcomes and technical developments which society would be better served keeping; the ones it should discard; and the ones which require further negotiation. OMDDAC findings are well-placed to inform this crucial phase, by combining insights from key literature, the views of a representative sample of the public, the experiences of key stakeholders and the voices of children and young people.

Methodology

OMDDAC is a collaborative and integrated interdisciplinary project. It includes researchers with expertise in law, governance, ethics, data and computer science, human-computer interaction, statistics, technology, and surveillance studies. It has adopted an innovative mixed-methods research design, incorporating case study analysis, stakeholder interviews,⁵ representative public surveys, and engagement with young people through a children's rights charity. Its outputs, including this project report, aim to apply project research findings to addressing policy and operational issues. For further details regarding our methodology, please refer to Appendix I.

⁴ UK National Data Strategy (n 1).

⁵ Being individuals with relevant expertise and/or experience in relation to the data-driven pandemic response. Participant stakeholders were drawn from a wide range of relevant sectors, backgrounds and disciplines, including: the private sector; government; academia; police and law enforcement; organisations specialising in the use, management and protection of data; the medical profession; charities and the third sector; regulators and the legal profession. See Appendix I.



Report Structure

This report brings the strands of OMDDAC's research together, combining and presenting the novel findings and recommendations from each method employed. In doing so, four key themes have been identified as underpinning our findings, which mirror the four pillars of the UK Government's National Data Strategy.⁶ The following four pillars, which will shape the future of data strategy in the coming years, have thus been applied to structure the contents of this report:

1. *Data foundations*
2. *Data skills*
3. *Data availability*
4. *Responsibility*

Defining 'Data-Driven'

During the landscape mapping assessment initially undertaken by OMDDAC,⁷ we identified that data-driven approaches to COVID-19 appeared to be understood by interviewees as existing 'on a spectrum between entirely automated, AI-powered processing and the more "mundane" uses of digital information and statistics' which feed into human decision-making.⁸ Applying this broad understanding of 'data-driven', the research team examined a range of areas where data use and public policy intersect to gain an understanding of how the public health crisis has prompted data-driven decision making:

- Data-driven public policy
- Tech-driven approaches to public health
- Policing and public safety
- Data in decision-making for children
- Public health surveillance of wastewater
- Public willingness to share data



⁶ UK National Data Strategy (n 1).

⁷ See Appendix I.

⁸ OMDDAC, 'Mapping the Data-Driven Landscape' (2021) <https://www.omddac.org.uk/news/mapping-the-data-drivenlandscape/> accessed 16 September 2021. In the context of our research, we limit AI to the use of data analytics and machinelearning algorithms.



1. Data Foundations

Interviewees consistently highlighted challenges relating to data quality and operational capability as hindering more effective data-driven responses to COVID-19. The importance of this issue is reflected in the first core pillar of the UK Government's National Data Strategy: Data Foundations. According to the Strategy:

'The true value of data can only be fully realised when it is fit for purpose, recorded in standardised formats on modern, future-proof systems and held in a condition that means it is findable, accessible, interoperable and reusable. By improving the quality of the data, we can use it more effectively, and drive better insights and outcomes from its use'.⁹

The National Data Strategy identifies obstacles to the effective sharing and use of data arising from the lack of consistency in the standards and data infrastructure used across the public sector. Our recommendations reflect these findings and address the associated challenges.

1.1 Data Quality

Stakeholders across the public sector consistently reported issues regarding data quality, referencing errors, omissions and out of date information. One of the central concerns of local authority stakeholders arising from this issue was that vulnerable residents who required support would be missed as a result, with data being received in incomplete (and as described by one stakeholder 'chaotic') forms:¹⁰

'We had examples where data received from NHS of persons who were on the shield list, where the address data was so bad that we only got a postcode. So, because this person may be in their house and not allowed or not able to leave, and maybe had no recourse to getting food or whatever, especially during the first lockdown, they were literally sending out people with nothing more than a postcode, trying to ask neighbours if they knew of this person to track them down (...) One high rise block of flats has one postcode so it could be hundreds of dwellings in some places. The importance of accurate addressing became really clear'.¹¹

Issues with data interoperability were also experienced, arising from inconsistencies with the data ontology and labelling used across different public sector organisations. As one interviewee explained: 'once you've agreed to share the data, bringing it together requires lots more engineering and familiarisation with the data than would be necessary if we had widespread data standards... Investment in our data assets is a big [challenge]. It is a little technical and boring, but if we've got good quality data that is well described and that has content standards around it, we can really accelerate the pace at which we can bring them together'.¹²

⁹ UK National Data Strategy (n 1).

¹⁰ Interview with government stakeholders (G12 and G13).

¹¹ Interview with government stakeholders (G12 and G13).

¹² Interview with government stakeholder (G17).

¹³ HMICFRS, 'Policing in the pandemic - The police response to the coronavirus pandemic during 2020', April 20 2021, p.16.

From the perspective of policing, data acquisition and data quality were two of the main challenges faced in monitoring resourcing and staff wellbeing during the pandemic. A report released by Her Majesty's Inspectorate of Constabulary and Fire and Rescue Services (HMICFRS) found that the IT systems in some police forces were detrimental to their understanding of demand for their services and how much resource they would need to commit to meet that demand. If the platforms for storing important data are themselves sub-standard, data interoperability within a force and across other forces is undermined, and detecting any missing data is made harder.¹⁴

OMDDAC research found that this had consequences for data-driven decision-making, highlighting the need for additional work to firstly identify uncertainty and incompleteness within the data and secondly to improve its quality. In the public health context, for example, the deployment of risk-scoring algorithms such as QCovid¹⁵ has emphasised the importance of raw data quality in the process of developing, testing, and validating an effective algorithm. One stakeholder highlighted that a key challenge presented by the QCovid algorithm, or indeed any algorithmically-informed decision-making tool, is the use of outdated or unreliable health data to inform risk stratification, emphasising the importance of steps to improve the quality of raw data:

*'To train a good machine learning algorithm you need lots of data and so we have places within the health system that will naturally have large quantities of data, but that doesn't necessarily mean it is clean data or well labelled data. And then when you use that data to train, you inherit both its clinical utility and the faults.'*¹⁶

Data quality is therefore essential in developing an algorithm that can be ethically and reliably used to inform clinical or policy decisions. Interviewees highlighted that where the quality and reliability of data is low, the robustness of automated, algorithmic decision-making is called into question, not least given the risk of 'spurious correlations' being formed.¹⁷ This underscores the continued importance of human oversight in the application of automated tools and AI which impact upon decisions affecting individuals; a feature which is particularly relevant in the context of the data protection reforms currently being proposed by the UK Government.¹⁸

Stakeholders additionally pointed out that there are a number of fundamental limitations inherent in the process of working with quantitative data, which must be taken into account and addressed transparently. As one stakeholder emphasised:

*'some things we are observing during the pandemic cannot be mathematically reduced into facts and requires some form of interpretation, coupled with local knowledge'*¹⁹

¹⁴ Relatedly, the National Law Enforcement Data Programme (NLEDS), a programme launched in 2016 to replace two police IT systems, has come under criticism by the National Audit Office due to delays which are, according to the NAO, putting police access to vital information at risk. The two systems (the Police National Computer and the Police National Database) are considered part of the critical infrastructure; however, their technology is becoming obsolete. The NLEDS programme, which had an original planned delivery date of 2020, was subject to independent review which found that if it was to continue as it was, it would be late, difficult and costly to roll out and maintain, and would not meet the needs of the police. National Audit Office, The National Law Enforcement Data Programme (HC 663 2021-22) <https://www.nao.org.uk/wpcontent/uploads/2021/09/The-National-Law-Enforcement-Data-Programme.pdf> accessed 17 September 2021.

¹⁵ See <https://www.qcovid.org/> accessed 16 September 2021.

¹⁶ Interview with data specialist stakeholder (D12).

¹⁷ Interview with data specialist stakeholder (D9).

¹⁸ The UK Government is seeking views on a proposal to remove Article 22 of the UK GDPR, with solely automated decisionmaking being permitted where it meets a lawful ground. DCMS, 'Data: a new direction' (10 September 2021) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1022315/Data_Reform_Consultation_Document_Accessible.pdf 1 October 2021.

¹⁹ Interview with regulatory stakeholder (R1).



Several stakeholders have addressed this contextual gap by supplementing their quantitative, administrative data with local knowledge and input, as well as qualitative research in the form of interviews and surveys. Local knowledge, in particular, proved to be a crucial element in informing appropriately targeted pandemic responses within the community.



Recommendations:

- **As part of the National Data Strategy, a national public sector data standards framework should be introduced and implemented.** This should incorporate clearly defined and consistent formatting, with labels and ontology being implemented across health and non-health authorities to enable interoperability. A rigorous 'lessons learned' exercise across authorities to inform such a framework will ensure best practice can be synthesised and deployed.
- **The National Data Strategy should also require that investment, resource, and corrective action be directed to formally address data quality issues across the public sector.** Getting the basics of data quality right must be a primary consideration before engaging in more sophisticated data analysis. Errors, omissions and outdated information should be highlighted for rectification. Collecting, cleaning, and harmonising data are important steps in the data 'lifecycle' which must be provided with sufficient time and resource.
- **Qualitative data should be formally collected and used in support of quantitative data to inform decision-making.** Quantitative datasets should be contextualised with qualitative findings from, for example, local knowledge or interview data.
- **Human oversight mechanisms should continue to play a central role in relation to the application of automated decision-making tools which feed into decisions affecting individuals.** This should consider whether the application of automated decision-making is appropriate given the particular context, taking into account the quality of the data available. Further, there is a primary need for human involvement to ensure data quality given the direct impact upon the robustness of outputs from automated methods of data analysis and decision-making.

1.2 Data Infrastructure

In terms of data infrastructure, stakeholders highlighted a huge diversity in technical and operational capability. Interviewees in local government, for instance, told us that while some authorities were able to operate using 'more sophisticated' systems, others relied heavily on software such as Microsoft Excel for data processing, limiting the opportunities for more advanced data analytics. The process for requesting and obtaining data from central government was also consistently described in onerous terms, requiring a significant element of manual processing on a daily basis. In addition, interviewees expressed concerns around data being shared through non-secure channels.

Similarly, data infrastructure posed various problems for a more coordinated data-driven response to the pandemic in the context of policing. One interviewee shared: 'There are at least 40 different approaches to tech and digital capability – different recording platforms even across different areas in a single force. That's why we have ended up using Excel'.²⁰ It was emphasised that the majority of data-driven responses did not involve very technically advanced analytics or machine learning. This raises two significant points: firstly, that valuable insights can be derived using already established data analysis techniques; and secondly, that greater consistency in data infrastructures would enable interoperability and could allow organisations to take advantage of more advanced tools, producing better outcomes across all forces.



Recommendation:

- **There should be significant consultation across the public sector, as part of the National Data Strategy to determine ICT infrastructure and resource requirements in order to facilitate appropriate and secure data sharing and analysis.** All public sector bodies should be involved in such a consultation, to improve levels of consistency in the data infrastructures used and enable more consistent methods of analysis and secure sharing of data.

²⁰ Interview with law enforcement stakeholder (L11).



2. Data Skills

Stakeholders told us that the pandemic has exposed several important skills gaps with regard to data-driven approaches, a theme which is reflected in the second core pillar of the National Data Strategy – Data Skills:

‘To make the best use of data, we must have a wealth of data skills to draw on. That means delivering the right skills through our education system, but also ensuring that people can continue to develop the data skills they need throughout their lives... To address this objective we will need to drive change across... capability, leadership and culture: developing world-leading capability in data and data science across central and local government, so that leaders understand its role, expert resource is widely available, staff at all levels have the skills they need, and a ‘data-sharing by default’ approach across government tackles the culture of risk aversion around data use and sharing’.²¹

In particular, our findings identified an urgent skills gap in relation to the data literacy capabilities of public sector leaders and decision-makers at a central and local level, which needs to be addressed given the importance of data for informed decision-making. Our recommendations set out to highlight and address these issues.

2.1 Data Literacy

Stakeholders identified a ‘breakdown at the interface between analysing data and deciding what to do with it’.²² In particular, interviewees referred to the inherent complexities of statistical modelling, and questioned whether policy- and decision-makers were properly equipped to make use of modelling outputs, leading to public concerns of the Government ‘hiding behind white coats for tough decision-making’.²⁴ As explored further in section 4.1 below, interviewees highlighted the importance of ensuring that modelling outputs were not presented, or perceived as, ‘the truth’; decision-makers must have a thorough understanding of the limitations, uncertainties and assumptions involved to be able to meaningfully interpret, evaluate and interrogate statistical outputs. In addition to the provision of training, therefore, meaningful, interdisciplinary collaboration between policy makers and specialists within a relevant domain was regarded by interviewees as crucial to ‘bridge the gap’²⁵ in understanding.

²¹ UK National Data Strategy (n 1).

²² Interview with academic stakeholder (A3).

²³ By statistical modelling, we refer here to data modelling as a probabilistic exercise focused on forecasting and predicting future outcomes.

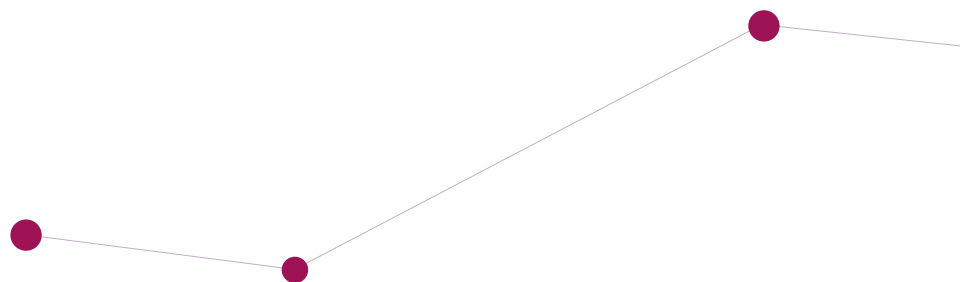
²⁴ Interview with government stakeholder (G3).

²⁵ Interview with government stakeholders (G9 and G10).

This data literacy skills gap, particularly at a leadership level, was also highlighted as a potential barrier to future progress regarding more innovative uses of data. As one stakeholder observed, 'our strategic director... understands the strategic need to make use of it, [but] is not very confident when he talks on the subject and that comes through'.²⁶ With regard to more advanced data science capabilities, the development of a 'larger workforce with AI expertise' forms part of the recently published National AI Strategy,²⁷ which acknowledges that there is a need for 'non-technical' employees to understand the 'opportunities, limitations and ethics of using AI'.²⁸ The establishment and incorporation of a professional framework for data science, including chartered status,²⁹ was also encouraged by interviewees as one approach to addressing the 'challenge, particularly in the public sector, that leadership doesn't have the expertise' in addition to the 'lack of a framework for what they are looking for in terms of skills' when hiring.³⁰

Relatedly, one of the biggest challenges for policing in relation to skills has been data interpretation. One stakeholder commented that 'sometimes I don't think we understand what's in front of us... support to interpret the data is not always available'.³¹ This is a particularly pressing concern given the vast quantities of data that police forces have been expected to sift through and derive meaning from – if there is an inability to turn 'data' into 'information' then the value of collecting and sharing large amounts of data is brought into question.

In some instances, OMDDAC research found that these gaps were filled by dedicated advisory groups. For example, academics were embedded into Police Scotland's OptICAL group to lead data analysis, check that powers were being used equitably and ensure that certain social groups or demographics were not being disproportionately impacted.³² One interviewee explained that the utility of this approach came from having dialogues with informed individuals who are not in police uniform and who can provide a realistic perspective on particular matters from outside the policing environment.³³ This creates a setting for constructive challenge which is not just beneficial for the police's skills, but for the academics who gain a better understanding of the context of police work during a public health emergency.



²⁶ Interview with government stakeholder (G11).

²⁷ Department for Business, Energy & Industrial Strategy; Department for Digital, Culture, Media & Sport, and Office for Artificial Intelligence (2021) 'National AI Strategy' <https://www.gov.uk/government/publications/national-ai-strategy/national-ai-strategy-html-version> accessed 24 September 2021.

²⁸ Ibid.

²⁹ Such as that proposed by the Royal Statistical Society: <https://rss.org.uk/news-publication/newspublications/2020/general-news/professional-standards-to-be-set-for-data-science/> accessed 16 September 2021.

³⁰ Interview with data specialist stakeholder (D4).

³¹ Interview with law enforcement stakeholder (L10).

³² Interview with law enforcement stakeholder (L9).

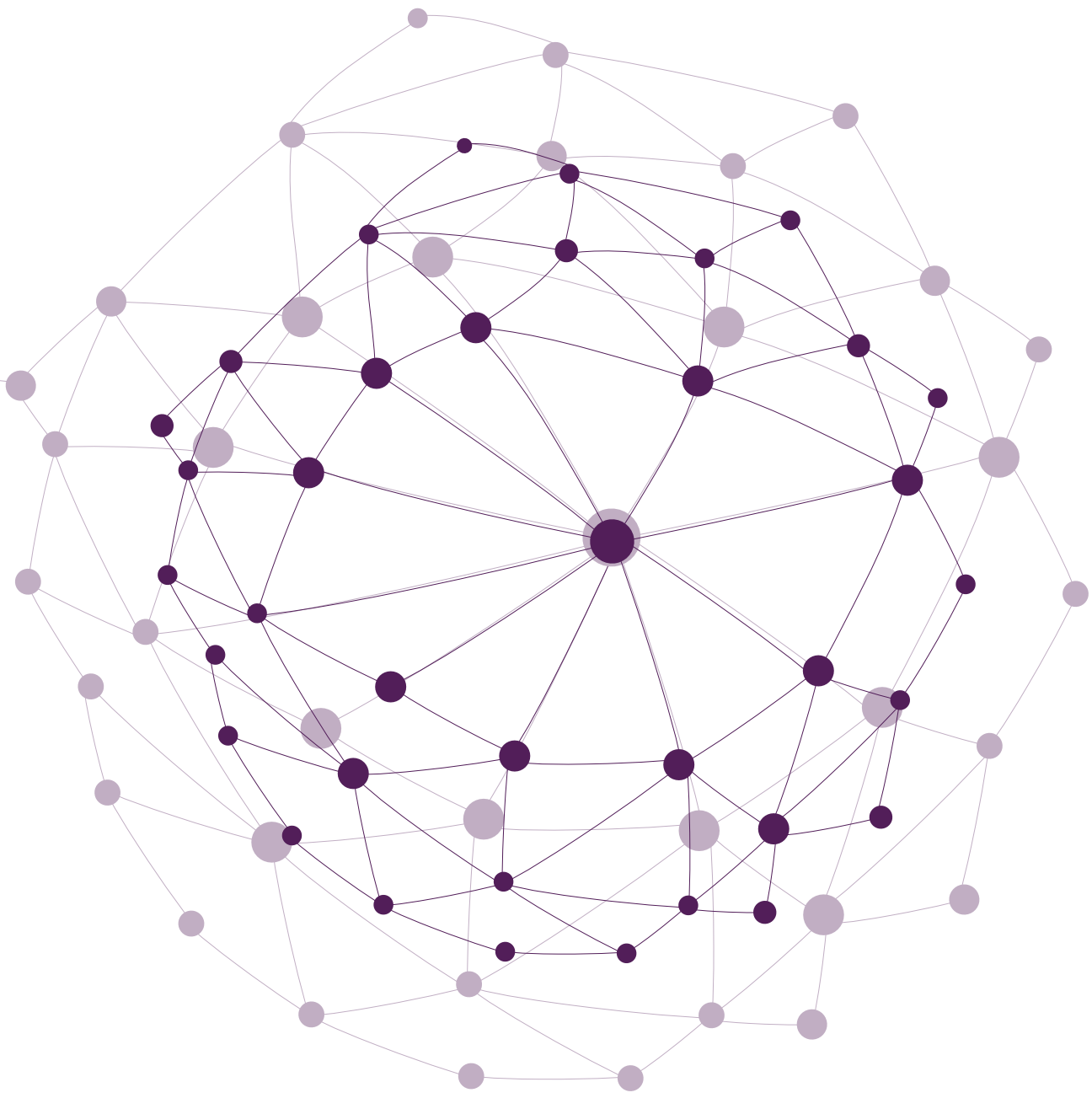
³³ Interview with law enforcement stakeholder (L12).





Recommendations:

- **Decision-makers and policymakers in the public sector should be required to undertake data literacy training.** Training should enable decision-makers to interpret, evaluate and interrogate statistical outputs and understand the associated risks and limitations. Professional standards for data science within the public sector should also be adopted to assist in this regard.
- **Meaningful multi-disciplinary collaboration should be actively encouraged.** Increased multi-disciplinary interaction is necessary to bridge the gap between statistical modelling and societal reality. More formal mechanisms are needed to allow specialists in the methodology (including epidemiologists, data scientists and mathematicians) to collaborate with domain specialists (including behavioural scientists as well as, critically, law



3. Data Availability

The sharing, linkage and repurposing of data sets from a diverse range of sources has been central to the national pandemic response. The third core pillar of the National Strategy - Data Availability - makes clear that this is a feature the Government intends to build upon:

'For data to have the most effective impact, it needs to be appropriately accessible, mobile and re-usable. That means encouraging better coordination, access to and sharing of data of appropriate quality between organisations in the public, private and third sectors'.³⁴

Our recommendations stress the continued importance of robust and visible information governance. Where necessary, the Government should be prepared to introduce statutory, legally enforceable limitations on itself and other public sector bodies which restrict how data is processed and shared in certain circumstances, to maintain an environment where data sharing and analysis is accepted and trusted by the public.

3.1 Data Sharing During the Pandemic

Existing and novel datasets have been combined in unprecedented, 'innovative' ways to better understand the prevalence of, and therefore respond to, COVID-19.³⁵ The types of data being shared include medical datasets as well as non-medical datasets such as mobility data.³⁶ Moreover, these datasets are acquired from both public and private sector sources (such as Google, retail, telecoms and social media companies).

Wastewater, for example, is one particularly novel data source which has been analysed to detect and quantify viral load in order to inform national and local pandemic responses.³⁷ The ability to identify traces of COVID-19 through data which is generally non-identifiable has been used to help provide an early warning system for new cases in a certain geographical area: 'The intriguing thing about wastewater is that it knows no statistical boundaries – the wastewater system reflects the way people actually live'.³⁸ Our research explored the perceptions of the general public, as well as younger members of the population, towards the sharing of wastewater data with different organisations, including local authorities, public health bodies, the police and a fictional commercial company. Our findings are presented in sections 3 and 4 of this report.

³⁴ UK National Data Strategy (n 1).

³⁵ See OMDDAC's Snapshot Reports for further in-depth analysis of data-driven responses to the pandemic: Rachel Allsopp and others, 'Snapshot Report 1: Data-Driven Public Policy' (2021) <https://www.omddac.org.uk/wp-content/uploads/2021/05/OMDDAC-Snapshot-Report-1-Public-Policy.pdf>; Sneha Dawda and others, 'Snapshot Report 2: Tech-Driven Approaches to Public Health' (2021) <https://www.omddac.org.uk/wp-content/uploads/2021/05/OMDDACSnapshot-Report-2-Tech-Driven-Approaches-to-Public-Health.pdf> and Ardi Janjeva and others, 'Snapshot Report 3: Policing and Public Safety' (2021) <https://www.omddac.org.uk/wp-content/uploads/2021/05/OMDDAC-Snapshot-Report-3-Policingand-Public-Safety.pdf> accessed 16 September 2021.

³⁶ For example: telecoms, transport, and geo-location data. Other non-health data sources include employee and school absence data.

³⁷ Defra, 'Tracking Covid-19 in wastewater' (2021) <https://deframedia.blog.gov.uk/2021/04/26/tracking-covid-19-in-wastewater/>; and Linda Geddes, 'We're the Poo Crew: Sleuths Test for Covid by Reading Signs in Sewage' *The Guardian*, <https://www.theguardian.com/world/2021/apr/24/were-the-poo-crew-sleuths-test-for-covid-by-reading-signs-in-sewage> (24 April 2021) both accessed 16 September 2021.

³⁸ Interview with government stakeholders (G9 and G10).



Within government, at both a central and local level, the increase in data sharing has assisted in numerous areas, including: supporting vulnerable residents within the community, predictive pandemic modelling and, more generally, informing public policy and pandemic-related responses. As one stakeholder shared:

'We've learned the importance of mobility data for helping with the pandemic response all the way across the board from modelling 'R', to understanding how to design lockdowns, what the parameters of those lockdowns are, the geographical areas for those locales... There is a sense of innovation there which is understanding the datasets that are designed for one purpose actually can give you insights in a whole different area... [which has been] really important in understanding non-observed pandemic spread'.³⁹

From the perspective of the police, opportunities for more effective data sharing were recognised both in terms of sharing data across policing stakeholders and receiving data from non-police stakeholders. With regard to the former, the centralised functions within the policing ecosystem such as the National Police Coordination Centre (NPoCC) quickly established mechanisms to monitor changes in key indicators like PPE quantities, staff sickness and absence rates, and COVID-19 rates in local police force areas. Forces were asked to provide daily returns on what their sickness levels were, how much of it was COVID-19 related, and how it was distributed across critical roles. This data would be collated and redistributed to forces to provide a longer-term picture of pressure on the policing service as a whole.

Whilst stakeholders were acutely aware of concerns relating to issues such as privacy and mission creep, as detailed in section 3.2 below, many of the public sector interviewees held the view that the more open approach to data sharing, which has been evident during the pandemic, has broadly been conducive to the public good, and therefore should continue in some form. The sharing of qualitative data, as discussed in Section 1, across different policing stakeholders was highlighted as an important part of their data-driven response. NPoCC, in partnership with the College of Policing, ran weekly online 'knowledge sharing events'. These covered a whole variety of topics, such as the latest trends on domestic abuse; the planned response to the re-opening of the night-time economy; how investigative techniques have been adapted; and how staff wellbeing is monitored. This allowed different forces to share their experiences on planning and lessons learned. One interviewee explained, 'it has given us a rich environment to say "Right, we understand some forces have experienced challenges with 'X', but this force has got a way of dealing with it"'.⁴⁰

3.2 Public Concerns

We conducted both qualitative and quantitative research analysing public concerns around the sharing of data for the purposes of the pandemic response.⁴¹ Our qualitative research involved semi-structured interviews with 70 stakeholders with experience or expertise in data-driven pandemic responses, whilst the perceptions of a representative sample of the general public in the UK towards data sharing were quantitatively analysed using two surveys.⁴²

³⁹ Interview with government stakeholders (G9 and G10).

⁴⁰ Interview with law enforcement stakeholder (L8).

⁴¹ All aspects of the OMDDAC study design were reviewed and approved by the ethical review process at Northumbria University.

⁴² See Appendix I for further detail.

We also engaged with 17 children and young people (aged between 11 and 18) with the support of children's human rights organisation, Investing in Children.⁴³ Children and young people have been described as the 'hidden victims' of the COVID-19 pandemic.⁴⁴ They have themselves said that they have not felt heard,⁴⁵ although many of the decisions that have been made during the pandemic have had significant impacts upon children and young people. Children have a right to have their views sought and due weight given on matters affecting them under Article 12 UN Convention on the Rights of the Child. Though the findings of this small-scale investigation into the views of children and young people cannot be considered representative of their views across the UK, they nonetheless provide an important insight into the opinions and concerns of some young people and identify key areas warranting further research.⁴⁶

Our research findings across these areas evidence that visible information governance controls are of central importance in creating an environment where data sharing will be accepted by the public.

Qualitative findings: Incorporating the perspectives of key stakeholders, including children and young people

Interviewees acknowledged public concerns around privacy, surveillance, and a risk of 'mission creep' in view of the amount and types of data being shared across public and private sector stakeholders. Our analysis of digital proximity and exposure notification apps,⁴⁷ for instance, highlighted the tension between expectations of the sharing of Test and Trace data and public understanding of how and where their data was being shared and used. The expectation of the NHS COVID-19 exposure notification app was that it would become a faster and more privacy-preserving method compared to human contact tracing. One interviewee stated that 'human contact tracing is much more invasive' and, in fact, needs to be as invasive as possible to be effective at managing the outbreak of the virus.⁴⁸ However, in order to ensure sufficient downloads of the app, the public required assurance that their data was sufficiently protected and would not be shared more widely. The same interviewee stated that 'the app could do much more if it did not have the data protection by design format' but this design was considered necessary to encourage sufficient uptake.

In general, the young people we engaged with understood the need to share information; all fifteen young people who expressed a view on this point said that they thought that it was acceptable, for example, for information to be shared with public bodies where COVID-19 was found in wastewater in a particular locality for the public good. There was disagreement, however, about who such information should be shared with (hospitals and doctors (12/15); local council (11/15); police (5/15); the Prime Minister (4/15)). If the information were to identify the young people and their families then only 9/15 participants would still consider such information sharing acceptable. One young person suggested that they would want to be made aware of such information sharing and be provided the opportunity to consent (or indeed refuse to consent).

⁴⁴ Barnardo's (n 3).

⁴⁵ Laura Lundy and others, 'Life Under Coronavirus: Children's Views on their Experiences of their Human Rights' [2021] 29 Int J Children's Rights 261-285.

⁴⁶ As explored throughout sections 3 and 4 of this report.

⁴⁷ See Sneha Dawda and others, 'Snapshot Report 2: Tech-Driven Approaches to Public Health' (2021)

<https://www.omddac.org.uk/wp-content/uploads/2021/05/OMDDAC-Snapshot-Report-2-Tech-Driven-Approaches-to-Public-Health.pdf> accessed 16 September 2021.

⁴⁸ Interview with academic stakeholder (A1).



Whilst stakeholders discussed an ‘institutional nervousness around sharing, linking and repurposing of data’⁴⁹ linked in the main to uncertainties regarding the application of data protection law historically, it was noted that some of the ‘cultural resistance has evaporated in the face of the pandemic’.⁵⁰ Consequently, stakeholders specifically emphasised that the importance of strong information governance becomes greater with increased data sharing.⁵¹ In this regard, stakeholders reported that they have been able to re-examine and operationalise existing data protection provisions for a more effective data-driven response. Guidance from the Information Commissioner’s Office on best practice in data sharing has been utilised, with one stakeholder reflecting that ‘it’s not the law that’s been the problem; it’s the issue of shared responsibilities... In ninety-nine percent of situations data protection law allows you to do what you want to do, provided the safeguards are there’.⁵² This sentiment was also recently reflected in the comments of the recently appointed Information Commissioner,⁵³ John Edwards, who rejected the notion of a dichotomy between data protection and data-sharing:⁵⁴

‘The UK DPA and UK GDPR are a “how-to” not a “don’t-do”. I think the UK and many jurisdictions have finally learned that lesson through the Covid crisis’.⁵⁵

Accordingly, data protection law can be interpreted as ‘enabling’ or facilitating the flow of information where necessary and proportionate, as opposed to simply being restrictive.⁵⁶ As such, information governance must play an ‘end-to-end’ role in all activities involving data, beginning with the initial data acquisition process. As part of this, to address public concerns, the implementation of privacy-preserving IT architecture and technology should always be considered as one possible option.

In addition, details of the information governance controls, mechanisms and restrictions on data sharing and processing must be made transparent to the public. This is necessary to provide visibility and reassurance to the public that the relevant risks have been considered and mitigated against, facilitating an environment where data sharing and analysis will be trusted.



⁴⁹ Interview with regulatory stakeholder (R4).

⁵⁰ Interview with data specialist stakeholder (D3).

⁵¹ By information governance, we refer here to the controls and processes in place for the management of data within an organisation pursuant to relevant legal obligations such as confidentiality and data protection.

⁵² Interview with regulatory stakeholders (R2 and R3).

⁵³ Department for Digital, Culture, Media and Sport (2020) ‘DCMS Committee approves appointment of John Edwards as next Information Commissioner’ <https://committees.parliament.uk/committee/378/digital-culture-media-and-sportcommittee/news/157409/dc-ms-committee-approves-appointment-of-john-edwards-as-next-information-commissioner/> accessed 15 September 2021.

⁵⁴ Digital, Culture, Media and Sport Committee, ‘Pre-appointment hearing for Information Commissioner’ (Thursday 9 September 2021) <https://parliamentlive.tv/event/index/1508ef32-ec1d-4295-a5bf-aa8cd8211e7a> accessed 15 September 2021 at 10:05.

⁵⁵ *ibid* 10:05-10:07.

⁵⁶ Interestingly, though data protection law was perceived by some stakeholders as a barrier to data sharing, this predominantly arose from uncertainty and a resultant risk-averse attitude, as opposed to the law being used as justification or an excuse to refuse to share data.

Quantitative findings: the public's perceptions of data sharing for COVID-19 purposes

We deployed two surveys to understand public perceptions of data-sharing for COVID-19 purposes.⁵⁷ The first survey explored the sharing of medical data and mobility data. The second explored the public's perceptions of wastewater sampling (a health surveillance approach that has been employed during the COVID-19 pandemic).⁵⁸

Both surveys were deployed using the software Qualtrics⁵⁹, and recruitment of nationally representative samples regarding age, sex, and ethnicity (according to the UK 2011 Census from the Office of National Statistics⁶⁰) was conducted by the intermediary Prolific.ac.⁶¹ 1,003 participants were recruited for the Medical and Mobility Data survey, and 299 participants were recruited for the Wastewater survey (see Table 1 in Appendix I for details of participant demographics).

Both surveys included questions that used a choice-based multiple conjoint structure, whereby participants were asked to compare pairs of data-sharing scenarios and for each pair choose the scenario that they preferred (see Figure 1 in Appendix I for an example of a paired-scenarios comparison). The data-sharing scenarios differed on a set of key data-sharing "attributes" (e.g. what data is being shared, and the UK COVID-19 Alert level under which the data-sharing request was made).⁶² By asking the participants to express their preferences over many pairs of different data-sharing scenarios, we were able to infer, via the use of multinomial logit regression,⁶³ the impacts of the attributes on data-sharing preference.

Inter-organisational data sharing preferences

Our results evidence that the public are likely to be concerned about data sharing between certain organisations. In both the surveys, we found that our participants preferred to share data with their Local Authority and preferred to share data with their devolved nation's Public Health Body even more, when compared to a commercial company, as shown in Figure 1 below. Also, our participants showed little preference for sharing with their Regional Police Force when compared to preference for sharing with a commercial company. These results were consistent across the Medical and Mobility Data survey and the Wastewater survey – despite the broadly non-identifiable nature of wastewater data - suggesting that the organisations that the UK public prefers to share their data with is consistent irrespective of the type of data is being shared.

⁵⁷ For more findings and recommendations from both surveys please see: Selina Sutton and others, "SNAPSHOT REPORT 4: Survey of Public Perceptions of Data Sharing for COVID-19 related purposes" (2021) <https://www.omddac.org.uk/news/snapshot-report-4-public-perceptions-survey/> and OMDDAC, 'Survey 2: Wastewater Health Surveillance' (2021) <https://www.omddac.org.uk/news/survey-2-wastewater-health-surveillance/> both accessed 5 October 2021.

⁵⁸ Someone who has COVID-19 will shed fragments of the virus into their faeces, which then enter the sewage system when they go to the toilet. Therefore, taking a sample of wastewater from the sewage system and analysing it to detect these virus fragments can indicate the presence and prevalence of COVID-19 outbreaks. For further information, see Matthew Wade and others, 'Wastewater COVID-19 Monitoring in the UK: Summary for SAGE – 19/11/20' (2020) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/940919/S0908_Wastewater_C19_monitoring_SAGE.pdf accessed 16 September 2021.

⁵⁹ Qualtrics, 'Qualtrics.' Version June 2021. <https://www.qualtrics.com> accessed 16 September 2021.

⁶⁰ Office of National Statistics, '2011 Census.' <https://www.ons.gov.uk/census/2011census> accessed 12th August 2021.

⁶¹ Prolific, 'Prolific.' Version June 2021. Available at: <https://www.prolific.co> accessed 16 September 2021.

⁶² See Tables 3 and 4 in Appendix I for more details on attributes and their settings.

⁶³ Kenneth E. Train, *Discrete Choice Methods with Simulation* (2nd edn, Cambridge University Press, 2009).



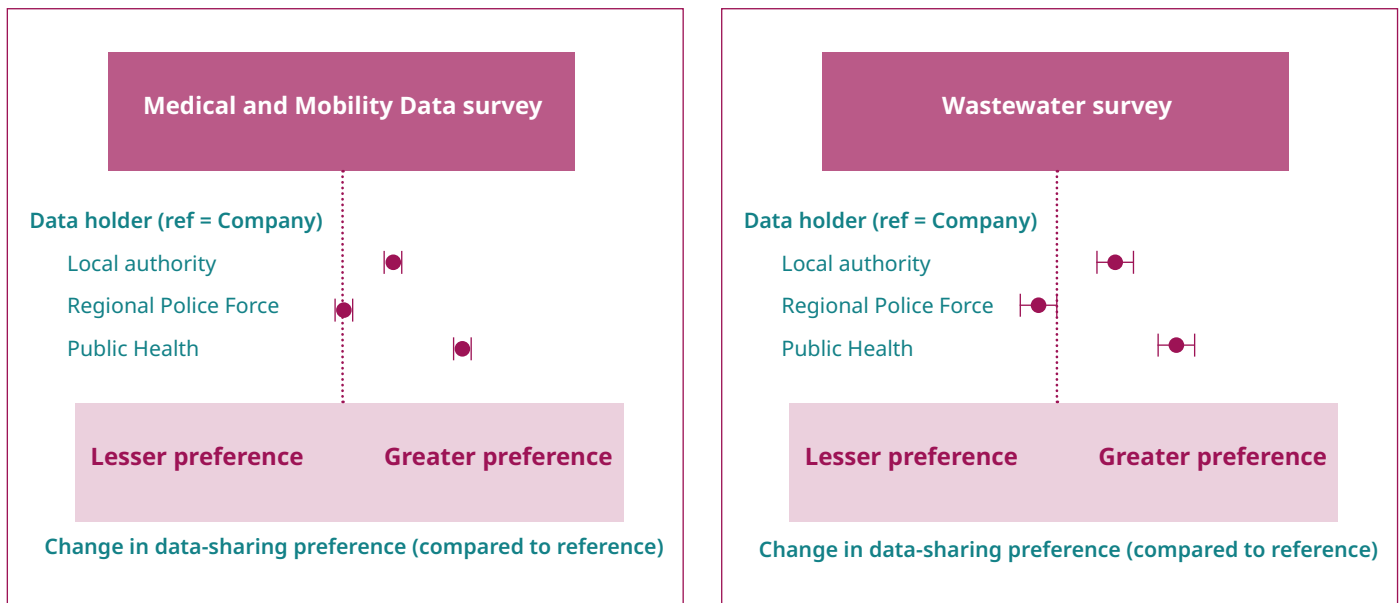


Figure 1. Change in preference when changing the data holder attribute (with all other attributes held fixed) from the Medical and Mobility Data survey (left) and the Wastewater survey (right). Point estimates (dots) and 95% confidence intervals (vertical bars) were obtained via multinomial logit models. The vertical dashed line in centre of each graph indicates preference for reference value (in this case 'Company'). Other results overlapping with this line indicate no change in preference to this reference value.

Patterns of willingness to share data

Both of our surveys have consistently illustrated the diverse positions that the participants took when it came to willingness to share data. After selecting the data-sharing scenario that they preferred in each paired-scenarios comparison, participants were then asked to rate how willing they would be to share their data in their preferred scenario along a Likert scale of 5 options, ranging from “Very Unwilling” to “Very Willing”. For each survey, we carried out a latent class analysis⁶⁴ on the Likert scale data gathered from the scenarios that the participants preferred. The analysis forms clusters (or groups) of participants, where participants within the same cluster gave similar answers to the Likert scale questions. After examining various models with different numbers of clusters, we identified seven separate clusters of participants in the Medical and Mobility Data survey and five clusters of participants in the Wastewater survey. The similarity in grouping between the two surveys is striking, despite their different foci (one on medical and mobility data and one on wastewater data).

In both surveys, we identified a cluster of participants that we named “The Committed to Sharing” (15.3%: 153/1003 in the Medical and Mobility data survey and 33.3%: 100/299 in the Wastewater data survey), who almost always answered “Very Willing” to share data under their preferred scenarios. Similarly, the cluster that we called “The Eager to Share” appeared in both surveys. Participants in this cluster (20.5%: 206/1003 and 31.3%: 94/299) almost always chose “Willing” under their preferred scenarios.

⁶⁴ Bridget E. Weller, Natasha K. Bowen, and Sarah J. Faubert, ‘Latent Class Analysis: A Guide to Best Practice.’ (2020) 46(4) J. Black Psychol, <https://doi.org/10.1177/0095798420930932>.

There were, however, participants who were more wary of sharing their data in both surveys. For example, in both surveys, we identified “The Undecideds” cluster, the participants in which often answered “Neither Willing nor Unwilling” (Figure 2). Only a small number of participants, 65 (6.5%) and 20 (6.7%) respectively, in the Medical-Mobility survey and the Wastewater survey belong to this cluster. We also identified “The Resistant to Sharing” cluster in both surveys (9.2%: 92/1003 and 6.7%: 20/299), whose answers were mostly on the “Unwilling” side of the Likert scale (Figure 3).

Finally, in the Medical and Mobility data survey there was a cluster that we named “The Opt-Outers” – participants who almost always answered “Very Unwilling”, although this only contained 30 (3%) of the 1003 participants in that survey.

“The Undecideds” cluster

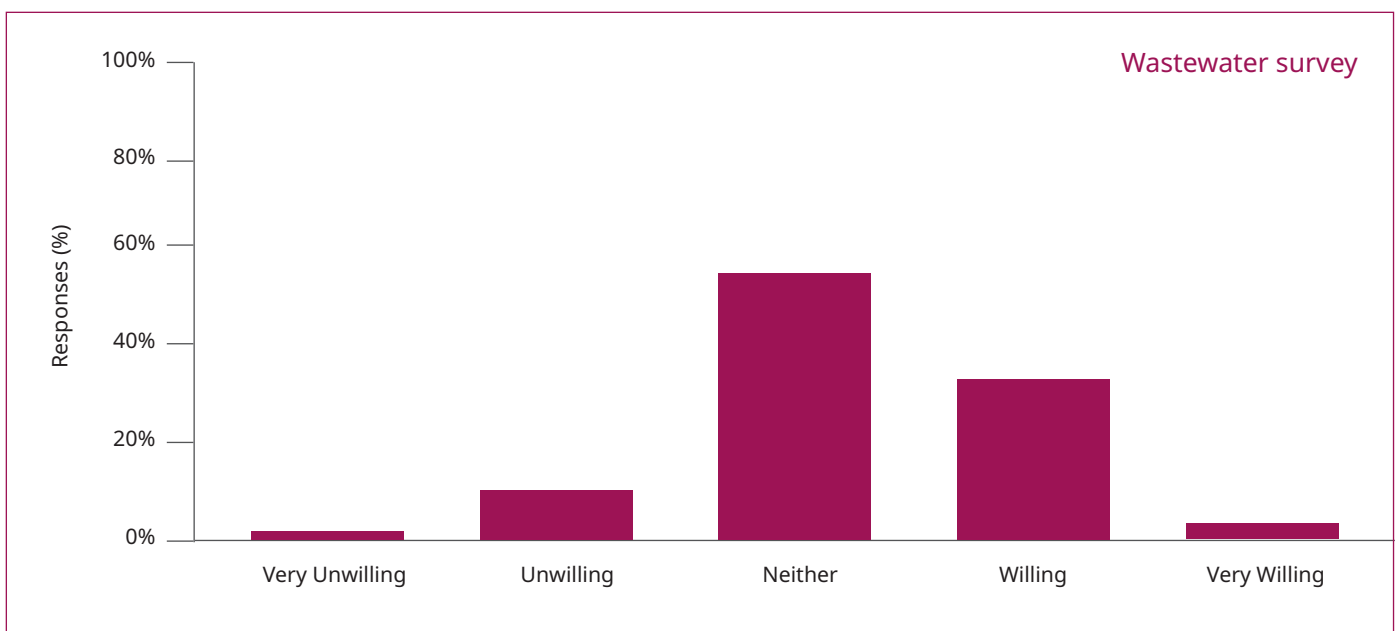
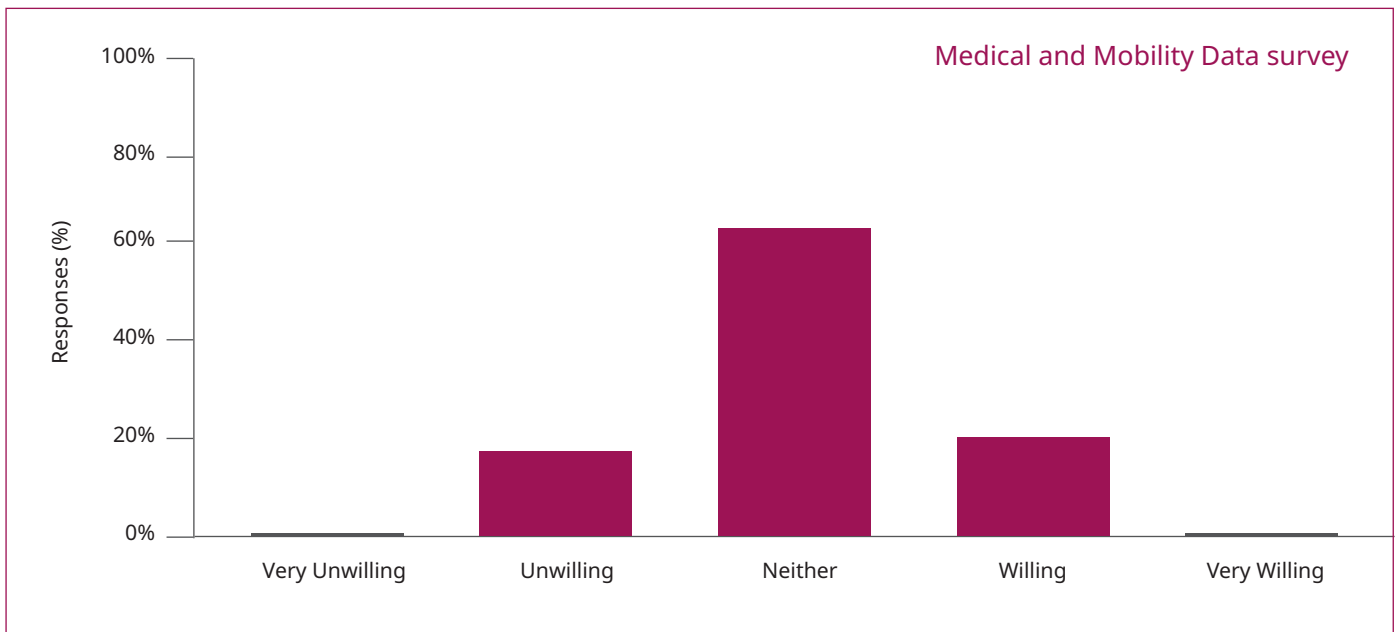


Figure 2. Patterns of data-sharing willingness for the participants in “The Undecideds” clusters. The taller the bar is, the more often the participants chose the corresponding option.



“The Resistant to Sharing” cluster

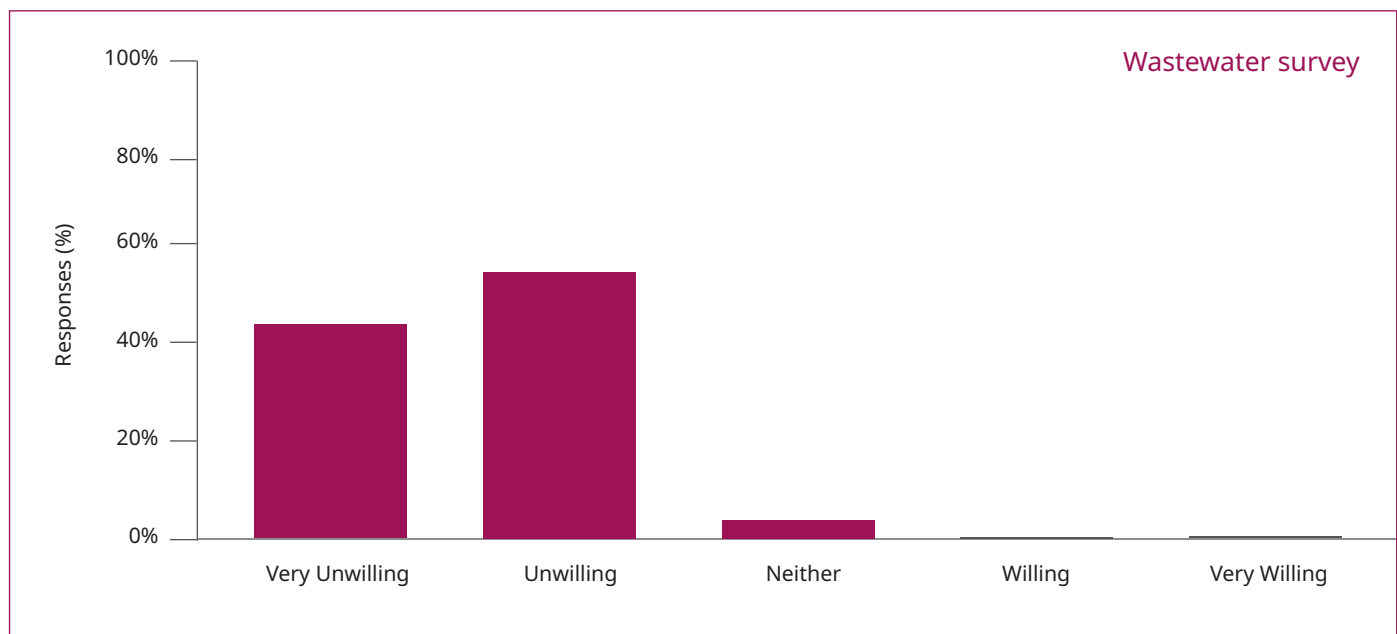
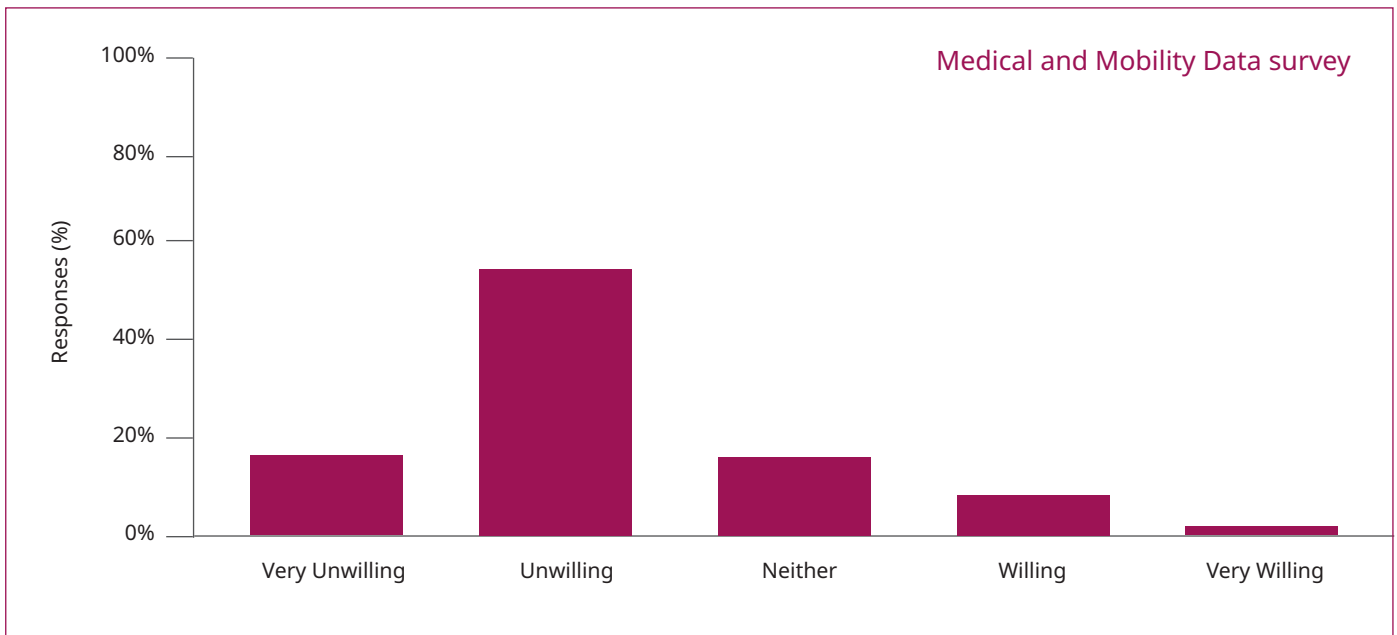


Figure 3. Patterns of data-sharing willingness for the participants in “The Resistant to Sharing” clusters. Same interpretation as in Figure 2.

This diversity of positions that our participants took further evidences the need for the information governance requirements behind data sharing to be accessible and transparent to the public, for it may be that through such communication, those who are currently unsure will have their concerns addressed. Of course, there may be members of the public that will not change their willingness to share data, but these approaches will still be beneficial in that such decision-making is likely to be more informed.



Data-detail preferences

Other findings from the Medical and Mobility Data survey evidence differences in what data our participants were willing to share with their preferred and non-preferred organisations. We devised two six-point scales, one for medical data and one for mobility data.⁶⁵ Starting with “No Data”, each option on these scales was a different type of data, their order reflecting increasing amounts of data detail. The final option on the medical scale was data from “Medical Records since birth” and on the mobility scale was data of “All destinations, and routes taken”. Participants were asked to choose what data, from the options along the scale, they would share with a particular organisation bearing in mind i) the UK COVID-19 Alert Level, and ii) whether the data was identifiable or ‘anonymous’. Therefore, to cover all the different combinations of Alert Level and ‘identifiable’ or ‘anonymous’, each of our 1,003 participants were asked this question 6 times for each organisation, resulting in 6,018 answers for each organisation.

The results of these questions underline the need for public transparency and consultation in inter-organisation data-sharing for there are notable differences in, for example, what data our participants would share with their Regional Police Force and their Public Health Body. When we consider the sharing of medical data across different contexts (different COVID-19 alert levels, and if the data was anonymous or identifiable), in most cases (37% of answers, 2,209 answers out of 6,018) our participants answered that they would limit the data they shared with their Regional Police Force to their COVID-19 test results, with another 31% of answers / 1,867 answers out of 6,018 saying that they would not give any data to their Regional Police Force. In comparison, in most cases (66% / 3,958 answers out of 6,018) our participants answered that they would be willing to share more detailed data than their COVID-19 test results with their devolved nation’s Public Health Body, including 23% / 1,414 answers out of 6,018 being willingness to share medical records since birth (see figure 4).

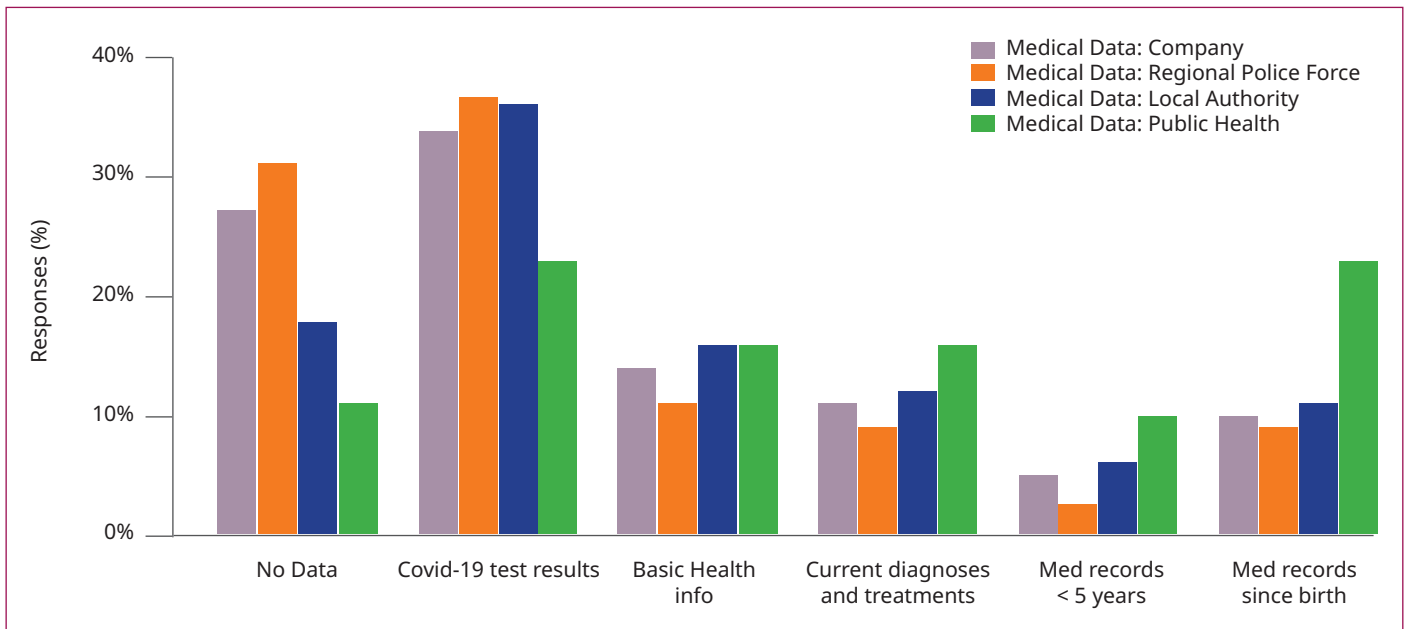


Figure 4. Results of the Medical Data Detail Slider questions by Data Holder

⁶⁵ See Appendix I for example questions.



However, our findings also indicate that there may be some types of data that the public would be less concerned about if it were to be shared across organisations. For example, when we consider the sharing of mobility data across different contexts (different COVID-19 alert levels, and if the data was anonymous or identifiable), the number of answers that indicate willingness to share what countries had been visited and when was relatively equitable across the four organisations (Company: 22% / 1,345, Regional Police Force: 20% / 1,199, Local Authority: 21% / 1,267, Public Health Body: 19% / 1,163 (all out of 6,018 answers)). The result was also similar for willingness to share which towns/cities had been visited and when (Company: 17% / 1,053, Regional Police Force: 18% / 1,083, Local Authority: 19% / 1,172, Public Health Body: 19% / 1,144 (all out of 6,018 answers)) (see figure 5).

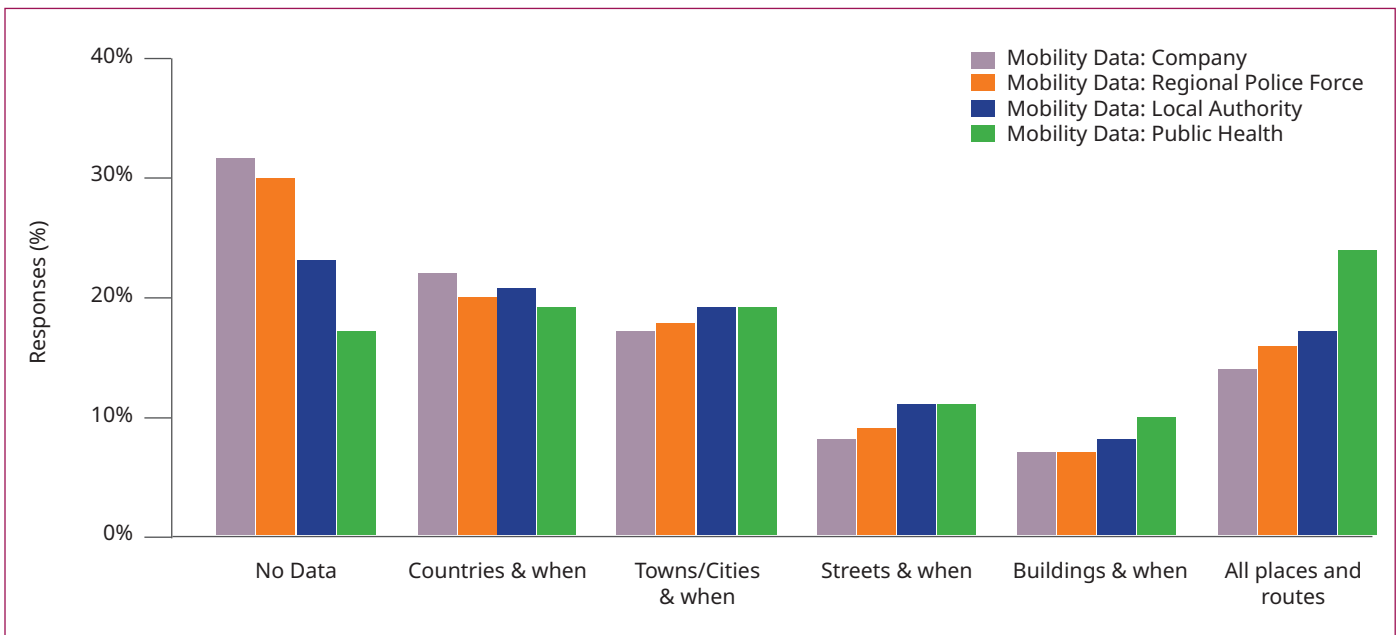
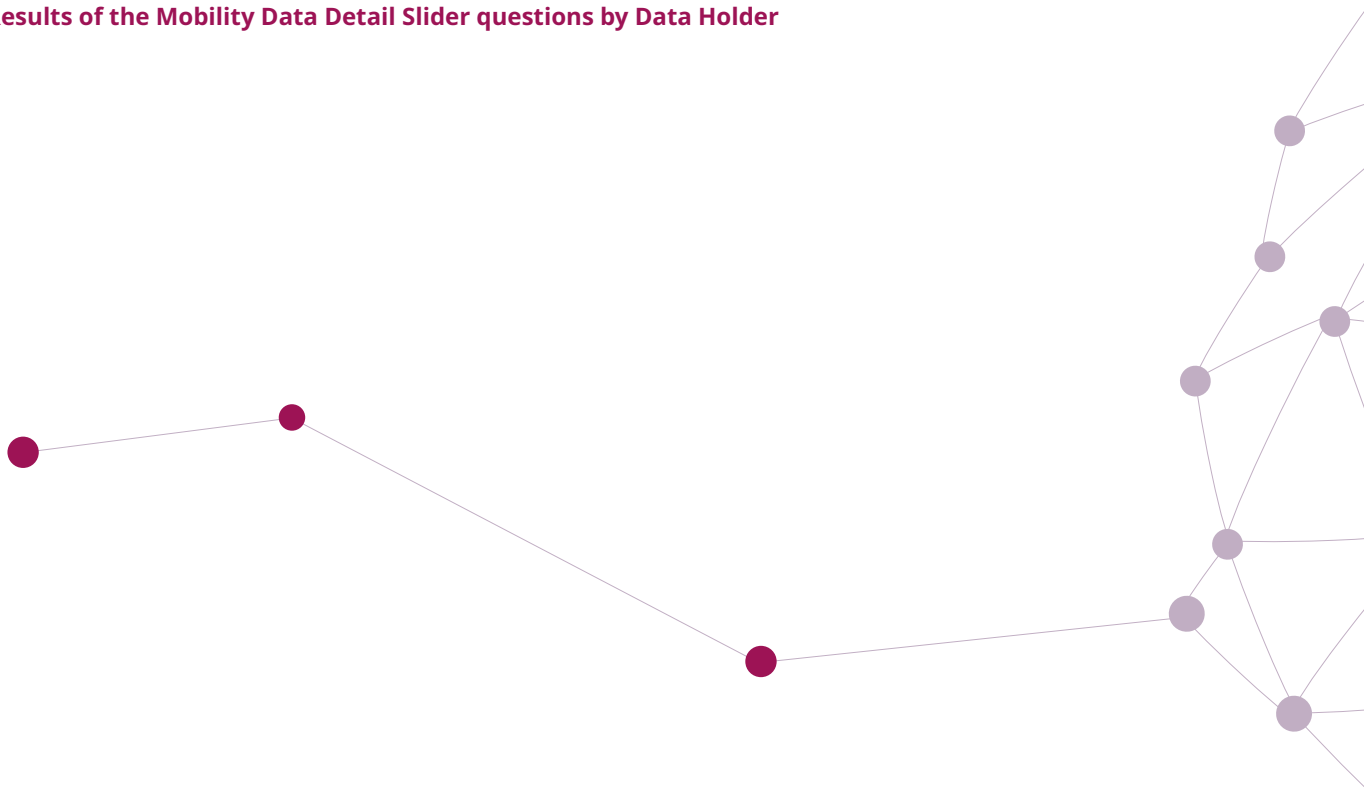


Figure 5. Results of the Mobility Data Detail Slider questions by Data Holder





Recommendations:

- **Data acquisition, sharing, and processing activities should begin and end with robust information governance requirements which are transparent to the public.** All data processing should undergo a data protection and human rights impact assessment at the outset. This assessment should not be regarded as simply a 'tick box' exercise, but instead should be a living document which is revisited regularly throughout the data 'lifecycle' on a continuous basis. Government and public bodies must also be prepared to impose legally enforceable limitations on themselves regarding the use of certain data. Information governance controls, including any such limitations, should be communicated to the public to create an environment where data sharing and analysis is trusted and accepted.
- **The design of any data-sharing initiatives should aim to foster and maximise public acceptability by:**
 - **being transparent about which organisations have access to certain types of data and when this data is being shared.** From the outset, individuals should have access to this information to enable properly informed decision-making. An information sharing notice should be published which explains what data is being shared and with whom.
 - **not assuming that the public are comfortable with their data being shared across all sections of the public sector.** This is irrespective of the type of data or the public purpose of the data use. This should be taken into account in determining the public acceptability of any data sharing initiative;
 - **avoiding a 'one-size-fits-all' approach.** Bespoke data-sharing initiatives and methods for communicating their design could be developed by understanding which social groups or demographics are most likely to either be dissuaded from, generally undecided about, or resistant to data-sharing and why, taking into account local knowledge and preferences;
 - **providing options for how much detail to share.** The ability to choose between several 'levels' of detail in the data that is being shared would allow individuals to engage with the data-sharing initiative in a way that they are comfortable with.



4. Responsibility

It is clear from our findings that all sharing and processing of data must be undertaken responsibly in order to preserve public trust. This key theme of 'responsibility' is reflected in the fourth core pillar of the National Data Strategy:

'As we drive increased use of data, we must ensure that it is used responsibly, in a way that is lawful, secure, fair, ethical, sustainable and accountable, while also supporting innovation and research'.⁶⁶

Our research findings identified a direct link between responsibility and the need for transparency regarding the use of data. This is also reflected in the National Data Strategy which makes clear that transparency is an important requirement for responsible data use: 'Enhancing individual awareness about the public benefits of data use requires transparency about such benefits, as well as a commitment to ensuring that people have trust and confidence in the use of their data and that it is adequately safeguarded'.⁶⁷ In our recommendations, we highlight the importance of transparency and public engagement regarding the way that data is used and shared, as well as the capabilities of data more fundamentally. This should be concerned with engaging all members of the population, including underrepresented communities as well as children and young people.

4.1 Transparency

The 'power' of data and its limitations

The issue of transparency was prominent throughout OMDDAC's study of data-driven responses to the pandemic. One essential aspect of this issue was in the use of the term 'data-driven' itself; interviewees raised concerns that by referring to a decision as having been 'driven' or 'informed' by data, it gives the impression that the decision is wholly objective or 'beyond dispute',⁶⁸ without giving due regard to the actual complexity of the situation.⁶⁹ As summarised by one stakeholder:

'the actual situation is much more complex. Those decisions are in turn informed by value judgements on which datasets are actually being used and whose data is actually being evaluated... we then make value judgements about what the data implies and what we are going to do with it. If we understand that facts are theory laden and value laden, then any use of the data to inform decisions implies values'.⁷⁰

Indeed, many of the decisions relating to the pandemic (for example opening schools or imposing lockdowns) have necessarily involved the weighing up of risks and relevant factors according to policy priorities: 'those decisions involve other aspects like impact on mental health and economic impact. These aspects are hard to compute, so the decision is complicated'.⁷¹

⁶⁶ UK National Data Strategy (n 1).

⁶⁷ Ibid.

⁶⁸ Interview with third sector stakeholder (T1).

⁶⁹ Relatedly, the Office for Statistics Regulation in its annual review report for casework between 1 April 2020 and 31 March 2021 highlighted that this year (2020/21) was the first year where the use or misuse of statistics was not the most common issue raised in casework; the category with the highest number of cases related to the quality and trustworthiness of data. The increase in this theme reflects the observation that the pandemic caused an increase in public interest in the trustworthy production and publication of statistics, particularly as statistics were being used so frequently and publicly to justify decisions affecting peoples' everyday lives (for example whether or not to impose restrictions): UKSA and OSR, Annual Review of UK Statistics Authority Casework 2020/21 (September 2021) <https://osr.statisticsauthority.gov.uk/wpcontent/uploads/2021/09/Annual-Review-of-UK-Statistics-Authority-Casework-2020-21.pdf> accessed 5 October 2021.

⁷⁰ Interview with academic stakeholder (A2).

⁷¹ Interviews with data specialist, government and academic stakeholders (G6, G9, G10, A8).

Moreover, while statistical modelling and data science have been instrumental in informing key pandemic decisions, stakeholders felt that greater transparency was required to ensure that outcomes of modelling were not presented, or perceived as, 'the truth'. Statistical modelling techniques are regarded as particularly effective in extracting additional value from data regarding potential future outcomes. Nonetheless, they inherently involve levels of uncertainty and rely on assumptions, for instance regarding behavioural parameters which cannot easily be measured.

I keep telling political leadership not to use maths beyond its power. People often ask you to solve a problem with modelling or statistics. Human judgement also needs to be involved. Maths can only tell you so much ... because we can't measure those behavioural parameters. ... it's modelling for illustration, modelling for building understanding, modelling for storytelling and that is it. Because that's such an important part of the epistemology of mathematical modelling and people often assume a naive "push button, computer tells answer" - this is not what I think the purpose of mathematical modelling is'.⁷²

Inclusive transparency: informing children and young people

More broadly, greater levels of transparency regarding how decisions have been reached has been called for as a matter of paramount importance given the unprecedented restrictions on liberty that have been introduced as part of the pandemic response.⁷³ Being open and transparent about the use of data is regarded as crucial to building public trust and aiding cooperation. According to the Public Administration and Constitutional Affairs Committee, 'this is not just a moral issue, it is integral to the response to this pandemic. Transparency builds trust and trust aids compliance with rules'.⁷⁴ Interviewees emphasised that this must be accessible for all sections of the population, including those from underrepresented communities as well as children and young people.

Whilst the young people who were engaged as part of our research were aware of some of the policies adopted by Government to tackle the pandemic, not all young people felt that they knew how Government is using information and technology to help tackle COVID-19.⁷⁵ Over half of the young people⁷⁶ who shared their views with us felt young people had not been given enough information about the Government's decisions during COVID-19, and a further three were unsure. There was concern expressed that action had not been taken to make young people more aware of data-driven responses to the pandemic, including for example the use of wastewater for predicting COVID-19 outbreaks. Moreover, seven of the fifteen young people shared that they were not aware of the proposed grading system for GCSEs and A levels in 2020,⁷⁷ three of whom were 17-18 years of age; a surprising finding given the significance of these proposals for young people.

⁷² Interview with data specialist stakeholder (D9).

⁷³ Public Administration and Constitutional Affairs Committee, Government transparency and accountability during Covid 19: The data underpinning decisions (HC 2021 803) <https://committees.parliament.uk/publications/5076/documents/50285/default/> accessed 16 September 2021.

⁷⁴ Ibid.

⁷⁵ In response to the questionnaire, three of the young people said they did not know how the Government is using information and technology to tackle COVID-19.

⁷⁶ 8 out of 15 young people.

⁷⁷ A-levels and GCSEs: How did the exam algorithm work? (BBC, 2020) <https://www.bbc.co.uk/news/explainers-53807730> accessed 24 September 2021.



Relatedly, the young people expressed concern at the lack of easily accessible, understandable and reliable information about the decisions that have been taken. To obtain the information they need to take informed decisions, the young people we engaged with tended to rely on schools, youth organisations, news broadcasts, their families and other trusted adults. Social media was also raised, though young people were critical about the reliability and, therefore, trustworthiness of information that was available on social media and suggested that more effective use could be made of government websites to provide information. Indeed, young people suggested these sites were often difficult to navigate, read and understand.



Recommendations:

- **To ensure transparency, future policy decisions informed by data modelling should be accompanied by explanatory statements which are published in a format that is accessible to all sections of the public.** The Government Office for Science should retain overall responsibility for this obligation, with each Chief Scientific Officer responsible for ensuring transparency of data use within their respective department. Statements should include:

 - Clear signposting on where and how to access the underlying data and statistics used to reach the decision (which should also be published);
 - Explanation of the analysis undertaken, accompanied by justification for the methodological choices used to inform the decision;
 - Details of relevant limitations or uncertainties associated with the data and/or analysis;
 - Details of assumptions which have been incorporated in the analysis;
 - Explanation of how factors which are not data-driven – including human judgements, behavioural considerations and policy priorities – have been incorporated into the decision.
 - A summary or digest of the above points, which conveys key findings, clearly avoids misrepresentation and explains what the findings do not show.
- **The general form and content that explanatory digests will take, as well as the means of communication, should be decided in consultation with different sections of the public, including young people.** Important information must be readily available through a variety of means to ensure sufficient access. Groups such as children and young people are not homogenous, meaning that multiple strategies are needed to reach them adequately.

Data sharing across the public sector

Our survey findings, as set out in detail in Section 3, highlighted the public importance of who data is being shared with. For example, our participants preferred scenarios where data was shared with their devolved nation's Public Health Body or their Local Authority as compared with their Regional Police Force. Furthermore, there was little difference in participants' preference for sharing data with their Regional Police Force, as compared with a commercial company.

Similar concerns were also expressed by the young people who engaged with our research. When asked whether it was appropriate for information to be shared regarding COVID-19 prevalence, all young people considered that data collected from testing wastewater should be shared in the public interest if anonymised. However, where that data is identifiable three of the young people expressed that they would not be willing to share, and a further two were unsure. Interestingly, whilst five of the young people stated that they would be happy to share wastewater data with the police, concern was expressed by others, due to a lack of clarity as to why the police would access to this information. Indeed, for some young people it was felt that 'the police had no business in knowing this information'.

In the context of policing, the issue of transparency was especially evident during our analysis of the Memorandum of Understanding (MoU) between the National Police Chiefs' Council and the Department for Health and Social Care, establishing a permissive regime for sharing of public health data. OMDDAC's research concluded that it was difficult for the general public to understand how public health data was being shared with the police and for what purposes.⁷⁸ Had the MoU been published earlier than it was, there could have been far greater confidence in the way that the police were accessing, handling, using and retaining the data under consideration. The delay allowed for uncertainty regarding the possible linkages of health data received by the police with entirely different records, and how the data is actively segregated from other data in police systems. This emphasises the importance not just of transparency, but the timing of it, and the manner in which key public concerns are addressed. In other instances, police forces in the UK have been more proactive in their approach to transparency. For example, one case study reviewed by OMDDAC was the creation of the Coronavirus Interventions System (CVI) by Police Scotland, a data recording tool that not only ensured up-to-date information on enforcement trends, but collated information on wider policing practice in relation to the other 'E's' in the police's '4E's' framework⁷⁹ for managing the pandemic. Data from the CVI was made available to produce valuable insights around the proportionality of enforcement. While this type of partnership certainly improved public scrutiny of how the police managed their response to the pandemic, it does not detract from the need to have formalised governance and oversight procedures in place at the same time.

⁷⁸ Ardi Janjeva and others, 'Snapshot Report 3: Policing and Public Safety (2021)' <https://www.omddac.org.uk/wpcontent/uploads/2021/05/OMD-DAC-Snapshot-Report-3-Policing-and-Public-Safety.pdf> accessed 16 September 2021.

⁷⁹ Enforcement, engagement with communities, explanation, and encouragement to return home. See: College of Policing, 'Understanding the law'. Available at: <https://www.college.police.uk/guidance/covid-19/understanding-law>.





Recommendations:

- All sections of the public, including young people, should be transparently informed about which organisations have access to their data.** Information must be made available to the public in an accessible form for all sections of the public in order to allay concerns about misuse and allow informed decision-making. Specifically, this information should be easy to find and understand. At a minimum, this should include what data is being collected, when, and if it will be stored with identifying information. Further, it should specify how and why an organisation is collecting, using and sharing data, and how it will be stored. In particular, details of inter-organisation data-sharing (e.g. a Public Health Body sharing data with the police), the limitations on such data-sharing, and the governance measures in place should be publicly available.
- Where Memoranda of Understanding are agreed between health agencies and enforcement bodies regarding sensitive health data, it is vital that the terms and conditions are made as transparent to the public as possible.** Transparency around the way that sensitive information is accessed, handled, used and retained should be the default position in similar situations in future. Clarifying the status of public health data received by the police in relation to other police databases is one important step in this.

4.2 Public Engagement and Trust

In general, interviewees were acutely aware of the privacy implications and negative public perceptions regarding the uses of data during the pandemic. Interestingly, however, it was reported by some local authority stakeholders that having consulted with members of the community there appeared to be a pre-existing assumption or expectation that data would be used in many of the ways that it has during the pandemic, often amongst young people. However, our research found that seven of the young people whose views were sought did not think that young people had been given enough information about the government's decisions during COVID-19, and a further three were unsure whether sufficient information had been provided.

For some young people, there appear to be public trust issues linked to a lack of discussion with them about how and why organisations might need to use young people's information; and more general concerns about certain organisations. Again, young people seemed to be particularly unclear about why the police would need information gained from wastewater analysis and expressed concern about the police having information about people being told to self-isolate. One participant raised the concern that the police could use that information simply to fine people, with a disproportionate impact on poorer families.

Ten of the young people whose views were sought considered that there had not been enough discussions with young people about the response to COVID-19, with a further ten expressing that young people's views should have been sought before decisions were made which impacted upon them, and their futures.⁸⁰ It should be acknowledged therefore that children and young people wish to express their views. They believe that they have much to contribute to discussion and that they should be consulted on issues which affect them. A more effective strategy is needed to inform and consult young people. The young people we engaged with suggested a range of different means by which government and policy makers can seek their views including: through schools, via surveys, through youth organisations and other trusted adults. The young people also suggested social media as another possible method of communication; however, they cautioned over-reliance on this method in light of its perceived lack of reliability and also, more pragmatically, because it cannot be assumed that all young people use social media. Multiple strategies for reaching children and young people should therefore be employed. Fundamentally, the most effective approach for reaching young people can only be determined by speaking to young people about how they want to be consulted and communicated with.

In a similar way to the discussion on transparency, our research found that there were elements of the police's data-driven response to the pandemic which risked undermining public trust. This was especially the case with the way in which the Amendment to the Health Protection (Coronavirus, Restrictions) (Self-Isolation) (England) Regulations 2020 (S.I. 2020/1045) was brought into law, without being debated in Parliament first. While it is difficult to quantify exactly how far this undermined public trust in the police's pandemic response, there is a risk that it could have perpetuated a view amongst sections of the public that the Government was treating the pandemic as a public order crisis rather than a public health crisis.

In general, government stakeholders that we spoke to were keen to emphasise the importance of public engagement going forward regarding future uses of data following the COVID-19 pandemic, so as not to make assumptions regarding public acceptability, to promote transparency and to ensure public confidence and trust.

'There could be implications for information governance post-pandemic. I'm careful about making assumptions, because we do have a particular licence to continue to do things safely, so it has been important that we don't omit that step because it's vital that we maintain public confidence in how we are using data both during the pandemic and afterwards. ... I think when looking to the future we're being very careful not to make assumptions about that public licence and need to make sure it's maintained; we are planning to do some further public engagement, looking carefully at what frameworks we can create that enable use of public data for innovation and creating public benefit. We are very much embarking on that conscious that we can't make assumptions that we just carry forward what's happening during exceptional times'.⁸¹

⁸⁰ A minority (4 young people) said they did not think young people should be asked before decisions were made which impacted on them (i.e. in relation to closing/opening schools, mask wearing at school etc.

⁸¹ Interview with government stakeholder (G17).





Recommendations:

- **The Government Office for Science should initiate a new public conversation post-pandemic, to consult with, inform and educate the public regarding the use of their data.** There are concerns amongst the general public over specific types of data-sharing which should be addressed promptly. A national communication campaign is needed regarding the use of medical/health data and the steps to be taken if people do not want their data to be used or shared in a certain way.
- **Public consultations must be inclusive of all members of the population, including underrepresented groups and young people.** Rather than a 'one-size-fits-all' approach, multiple strategies are required for consulting and communicating with different members of the population. These strategies should be determined by direct engagement with members of these groups, including children and young people.



Concluding Recommendation and Next Steps

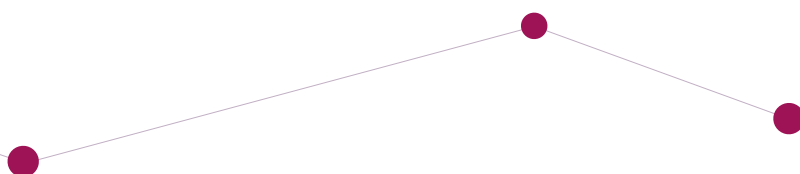
Our findings demonstrate the complexity of issues involved in the use of data for decision-making as well as the sharing of individual's data, particularly across multiple public sector authorities. The central importance of data quality and integrity, robust information governance mechanisms and public transparency for creating an environment where data analysis and sharing can be trusted and accepted in an emergency context is clearly apparent from our findings. In conclusion, we suggest that these obligations should be reinforced by a regulatory oversight mechanism:



Recommendation:

- **During emergency situations, an appropriate independent oversight body should be appointed** (with concomitant resourcing), with the remit to make sure that:
 - use and sharing of data by central government and public bodies for operational and public policy purposes serve the public good;
 - data use and sharing are supported by appropriate public transparency about data sources and methods of analysis;
 - data analysis is conducted with quality and integrity;
 - data sharing is overseen by independent governance and is subject to appropriate controls, conditions and time-limitations.

This function would be supported by a standing committee of senior representatives of public sector data controllers (reflecting national, regional and local bodies) in order to facilitate appropriate and time-limited arrangements for data sharing and analysis.



Further Research

In order to gain a greater understanding of the concerns of people who were either unwilling or very unwilling to share their data, a follow up survey should be undertaken to investigate the reasoning and factors underlying the public's perceptions and areas of uncertainty. In particular, there is a clear need for a deeper understanding of why public preference differs between mobility data and medical data, as well as the low preference with regard to sharing data with the police service, as evidenced in our survey findings. Further research to attain such an understanding will be important to inform future government data-sharing initiatives as well as the imminent data protection reforms.

Additionally, OMDDAC's engagement with children and young people highlights the need for further in-depth research in this area. The young people involved in our research made clear that they thought young people in general have not been considered enough during the pandemic. Our findings indicate that more could be done to ensure young people are able to access information about policies which impact directly upon them and, crucially, to capture and consider their views. It should be recognised that qualitative research exploring young people's views appears to have been 'thin on the ground'.⁸² OMDDAC's research illustrates that further investigation is necessary to better understand the views and opinions of children and young people towards the data-driven pandemic responses and, more fundamentally, to ascertain the most effective methods of engagement in order to ensure that their voices are heard.

Reflections and Learnings from a Distributed, Remote, Interdisciplinary Research Project

Finally, we reflect on the project's methodological approach. OMDDAC's interdisciplinarity has not been limited to a series of parallel single-discipline work-packages. The key to OMDDAC's methodology is the involvement of all researchers in the planning and conduct of the research, and production of outputs, with each output being led by an allocated researcher or team of researchers.

The project has been carried out in a fully distributed and remote manner. This experience has enabled us to identify benefits, risks and challenges which, if future remote and distributed research is to be successful, must be addressed.⁸³

⁸² Stephen Coleman and others, 'THE INTERNET ON OUR OWN TERMS: How Children and Young People Deliberated About Their Digital Rights' (2017) <https://casma.wp.horizon.ac.uk/wp-content/uploads/2016/08/Internet-On-Our-Own-Terms.pdf> accessed 4 October 2021.

⁸³ Rachel Allsopp, Claire Bessant, Keith Ditcham, Ardi Janjeva, Guangquan Li, Marion Oswald and Mark Warner, 'Observing data-driven approaches to Covid-19: Reflections from a distributed, remote, interdisciplinary research project' *Journal of Legal Research Methodology*, 1(1) <https://www.northumbriajournals.co.uk/index.php/jlrm/article/view/1160>.

These include:

1. The theoretical benefits often posited for interdisciplinarity research will only be realised by designing the integration – in terms of responsibilities and involvement of each discipline and the degree of interdisciplinarity expected - into every aspect of a project. There is a need for new university processes to support interdisciplinary researchers and cut through departmental and disciplinary barriers, with such integrated interdisciplinarity being sought out by funders in their calls.
2. Successful interdisciplinary collaboration depends upon building up understanding between disciplines. This requires time to build relationships, trust and understanding between team members using 'both task talk and relational talk.'⁸⁴
3. The challenges of communication and understanding that integrated interdisciplinarity creates can be transformative. Researchers are challenged 'to think hard(er) for effective communication and opportunity for the idea to develop further following the questions asked and contributions made by other researchers from different angles' leading to more insightful and impactful outcomes.⁸⁵
4. The interdisciplinary 'pair' online interviewing method developed by OMDDAC has expanded the range of individuals and organisations that the project has been able to reach, with considerable advantages from an inclusivity and diversity perspective. Planning for shorter interviews is necessary however to account for limitations on rapport-building and screen fatigue.
5. It will be a false economy for universities and funders to assume that research can continue to be conducted in a fully remote manner with consequent budget savings. There are certainly a number of opportunities for research presented by remote methodologies. It will important however for new researchers in particular that they are not isolated from opportunities to build working relationships and academic networks by in-person contact. Our research has indicated that certain groups such as young people may be particularly difficult to engage with by purely remote means.

⁸⁴ Jessica Leigh Thompson (2009) Building Collective Communication Competence in Interdisciplinary Research Teams, *Journal of Applied Communication Research*, 37:3, 278-297, DOI: [10.1080/00909880903025911](https://doi.org/10.1080/00909880903025911).

⁸⁵ 'Observing data-driven approaches to Covid-19: Reflections from a distributed, remote, interdisciplinary research project' (n 83).



Appendix I - Methodology

This appendix provides detail of the individual methods employed during the project. The OMDDAC qualitative and quantitative study design was reviewed and approved by the ethical review process at Northumbria University.

Stakeholder Interviews and Case Study Analysis

The OMDDAC project began with a foundational 'stakeholder mapping and landscape assessment', which was conducted between November 2020 and January 2021. Drawing on the multidisciplinary expertise and networks of the OMDDAC team, this was achieved primarily through document review, supplemented by remote interviews with 34 stakeholders, selected using a purposive, selective sampling strategy. Participant stakeholders were drawn from a wide range of relevant sectors, backgrounds and disciplines, including: the private sector; government; academia; police and law enforcement; organisations specialising in the use, management and protection of data; the medical profession; charities and the third sector; regulators and the legal profession. To maintain the anonymity of interviewees, interview notes were pseudonymised and all findings are presented in a way which is non-attributable.

Each interview during this initial research phase was conducted by two OMDDAC researchers, based upon a semi-structured interview guide, and typically lasted between 30 minutes and one hour. Questioning was led by one interviewer whilst the second focused primarily on contemporaneous notetaking, though follow up questions were also raised by the second researcher who often had a different disciplinary background or research interest. As a result, the breadth and depth of interview data collected was extensive. This 'pair-interviewing' approach also aided in ensuring adherence to the short timescales of this time-critical COVID-19 project, whilst maintaining the necessary rapport between the researchers and interviewees.

From this initial mapping exercise, specific case studies were identified (based on significance and potential impact) for deeper analysis during the second stage of the project, across three broad sectors: data-driven pandemic public policy,⁸⁶ tech-driven approaches to public health,⁸⁷ and policing and public safety.⁸⁸ Within these broad sectors, specific data-driven pandemic responses were identified as case studies for further investigation, as follows:

- Data-sharing, linkage and 'innovation';
- Predictive pandemic modelling;
- Data-driven support for vulnerable residents;
- Digital proximity and exposure notification;
- Risk-scoring algorithms;
- COVID-status certification;
- Police use of public health data;
- Monitoring of crime and enforcement trends; and
- Monitoring of police resourcing and wellbeing.

⁸⁶ Rachel Allsopp and others, 'Snapshot Report 1: Data-Driven Public Policy' (2021) <https://www.omddac.org.uk/wpcontent/uploads/2021/05/OMDDAC-Snapshot-Report-1-Public-Policy.pdf> accessed 16 September 2021.

⁸⁷ Sneha Dawda and others, 'Snapshot Report 2: Tech-Driven Approaches to Public Health' (2021) <https://www.omddac.org.uk/wp-content/uploads/2021/05/OMDDAC-Snapshot-Report-2-Tech-Driven-Approaches-to-Public-Health.pdf> accessed 16 September 2021.

⁸⁸ Ardi Janjeva and others, 'Snapshot Report 3: Policing and Public Safety' (2021) <https://www.omddac.org.uk/wpcontent/uploads/2021/05/OMDDAC-Snapshot-Report-3-Policing-and-Public-Safety.pdf> accessed 16 September 2021.

A purposive, selective sampling strategy was then used to identify participant stakeholders able to provide detailed information with regard to the above areas of concern. A semi-structured interview guide was collaboratively developed, enabling OMDDAC researchers from multiple disciplines to obtain data concurrently in a consistent manner. This approach, in addition, allowed for flexibility to probe specialised areas of knowledge and experience. Stakeholder interviews were conducted with 36 participants between February and April 2021, again remotely in pairs, and typically lasted approximately one hour. These interviews were also conducted on the basis that the identities of the interviewees would remain confidential, with their contributions non-attributable; this element of the research design being crucial to ensuring stakeholders felt able to freely share details of their pandemic experiences.

Analysis of the interview data was conducted using an inductive, grounded theory approach to identify codes representing recurring themes. On identification of such themes, further granular analysis was conducted to explore the cross-cutting trends and challenges experienced by stakeholders in greater detail, in order to identify the important lessons learned as set out in this research compendium.

Representative Public Surveys

In tandem with the interview data collection, we developed two surveys to explore the public's willingness to share their data for COVID-19 related purposes. The first survey explored two types of data that we as a nation have become somewhat familiar with sharing over the course of the pandemic: medical data and mobility data. The topic of the second survey was more specific. Someone who has COVID-19 will shed fragments of the virus into their faeces, which then enter the sewage system when they go to the toilet. Therefore, taking a sample of wastewater from the sewage system and analysing it to detect these fragments of the virus can indicate the presence and prevalence of COVID-19 outbreaks.⁸⁹ A particularly novel public health surveillance approach is wastewater sampling. Thus, we wished to explore the public's perceptions of this public health surveillance technique. Both the surveys were deployed using the software Qualtrics, and recruitment of nationally representative samples in regard to age, sex, and ethnicity (according to the UK 2011 Census from the Office of National Statistics⁹¹) was conducted by the intermediary, Prolific.ac.⁹² 1,003 participants were recruited for the Medical and Mobility Data survey, and 299 participants were recruited for the Wastewater survey (see Table 1 for further details).

⁸⁹ Matthew Wade and others, 'Wastewater COVID-19 Monitoring in the UK: Summary for SAGE – 19/11/20' (2020) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/940919/S0908_Wastewater_C19_monitoring_SAGE.pdf accessed 16 September 2021.

⁹⁰ Qualtrics, 'Qualtrics.' Version June 2021. <https://www.qualtrics.com> accessed 16 September 2021.

⁹¹ Office of National Statistics, '2011 Census.' <https://www.ons.gov.uk/census/2011census> accessed 12th August 2021.

⁹² Prolific, 'Prolific.' Version June 2021. Available at: <https://www.prolific.co> accessed 16 September 2021.



	Medical and Mobility Data survey	Wastewater survey
Age		
18-27	14% / 138	18% / 53
28-37	18% / 179	18% / 54
38-47	19% / 193	18% / 55
48-57	15% / 151	16% / 49
58+	34% / 340	30% / 88
Ethnicity		
Asian	8% / 76	10% / 29
Black	4% / 40	7% / 20
Mixed	3% / 26	4% / 12
Other	2% / 20	3% / 10
White	83% / 839	76% / 228
Sex		
Male	49% / 486	49% / 145
Female	51% / 514	51% / 153

Table 1. Participant demographics

To ensure good quality data, each survey included two sets of three attention checks, and participants were informed that there would be attention checks and that their data would be rejected if they failed some of them at the start of the survey. Table 2 details the number of submissions that were rejected as well as how many submissions were returned (the participant began the survey but then decided to leave), and timed-out (the participant became inactive).

	Medical and Mobility Data survey	Wastewater survey
Approved	1,003	299
Rejected	18	11
Returned	83	32
Timed-Out	18	13

Table 2. Survey submissions



For taking part in the Medical and Mobility Data survey, which took 17 minutes and 30 seconds on average to complete, participants received £2.23 in remuneration. The survey was released at 11:00am on 14 June and it took just over 48 hours to receive responses from 1,003 participants, the vast majority of whom submitted responses in the first 24 hours.

For taking part in the Wastewater survey, which took 13 minutes and 43 seconds on average to complete, participants received £1.37 in remuneration. The survey was released at 1:45pm on 3 September and it took just over 48 hours to receive responses from 299 participants, the vast majority of whom submitted responses in the first 24 hours.

Regarding question types, two were consistent across the two surveys, and one additional question type was asked in the Medical and Mobility Data survey.

In both surveys we asked participants to compare a series of data sharing scenarios and then choose the one which they preferred (see figure 1 for an example question). The ways in which the data sharing options differ are known as 'attributes' and the different options within each attributes are known as 'levels' (see tables 3 and 4 for the attributes and levels used in each survey). This is a Choice-based Multiple Conjoint Analysis study design. Two or more options that differ in multiple ways are compared side-by-side and the participant chooses their preferred option. It is thought that by forcing the participant to make a choice, aspects that are the most pertinent to such decision-making become evident. Therefore, this method deconstructs an overall judgement into its component parts and in doing so, the aspects that have the greatest relative influence on the decision can be identified.

In which of these two scenarios are you more willing to share your data? (19 questions remaining)

Scenario A		Scenario B	
We are in	ALERT LEVEL 5	We are in	ALERT LEVEL 3
You are sharing	<u>Medical Data</u>	You are sharing	<u>Mobility Data</u>
with	<u>Info-Insights</u>	with	<u>your Local Authority</u>
Your data is	<u>Identifiable</u>	Your data is	<u>Anonymous</u>
and used to monitor if the current alert level, and its associated restrictions, is appropriate.		and used to monitor if the current alert level, and its associated restrictions, is appropriate.	

Figure 1.* An example question from the Medical and Mobility Data survey. The attributes were i) the UK COVID-19 Alert Level, ii) Data Type, iii) the organisation the data is being shared with, and vi) whether the data is identifiable or anonymous.

* The visual design of the question used in the study differs to the one shown here. Please refer to our WP3 snapshot report to view the actual design (<https://www.omddac.org.uk/wp-content/uploads/2021/08/WP3-Snapshot.pdf>).



Attributes	Levels	Description
UK COVID-19 Alert Level	Alert Level 5	Extremely strict social distancing. Risk of NHS services being overwhelmed.
	Alert Level 3	Virus is in circulation. Some social distancing restrictions.
	Alert Level 1	Covid-19 no longer present in the UK. No social distancing restrictions.
Data Type	Medical	Health data gathered when using any NHS service.
	Mobility	Where you have been and when.
Data holder	Info-Insights (fictional company)	A company that has been hired to collect, store, and analyse data.
	Regional Police Force	Provides policing services in your area.
	Local Authority	Lowest tier of UK government. Provides public services, e.g. waste collection.
	[Public Health body]	Aims to make the public healthier through education and helping plan NHS services.
Storage	Identifiable	Your name and address are included.
	Anonymous	The data does not include your name or address.

Table 3. The attributes and levels used in the Medical and Mobility Data survey



Attributes	Levels	Description
UK COVID-19 Alert Level	Alert Level 5	Extremely strict social distancing. Risk of NHS services being overwhelmed.
	Alert Level 3	Virus is in circulation. Some social distancing restrictions.
	Alert Level 1	Covid-19 no longer present in the UK. No social distancing restrictions.
Sample Size	Street Level	Sample includes wastewater from approx. 100 people living on the same street.
	Estate Level	Sample includes wastewater from approx. 1,000 people living on the same estate.
	Town-City Level	Sample includes wastewater from approx. 100,000 people living in the same town/city.
	County Level	Sample includes wastewater from approx. 1,000,000 people living in the same county.
Data holder	Info-Insights (fictional company)	A company that has been hired to collect, store, and analyse data.
	Regional Police Force	Provides policing services in your area.
	Local Authority	Lowest tier of UK government. Provides public services, e.g. waste collection.
	[Public Health body]	Aims to make the public healthier through education and helping plan NHS services.

Table 4. The attributes and levels used in the Wastewater survey

After selecting the data-sharing scenario that they preferred in the Choice-based Multiple Conjoint Analysis study design question, participants were then asked to rate how willing they would be to share their data in their preferred and I their non-preferred scenario along a Likert scale of 5 options, ranging from “Very unwilling” to “Very willing” (see figure 2).





Figure 2.* Willingness Likert scale

The final question type was only used in the Medical and Mobility Data survey. We devised two six-point scales, one for medical data and one for mobility data (see figures 3 and 4). Each option on these scales was a different type of data, with their ordering reflecting different amounts of data detail they would be likely to contain, including a “No data” option. Participants were asked to choose what data, from the options along the scale, they would share with a particular organisation bearing in mind the i) UK COVID-19 Alert Level, and ii) whether the data was identifiable or anonymous.

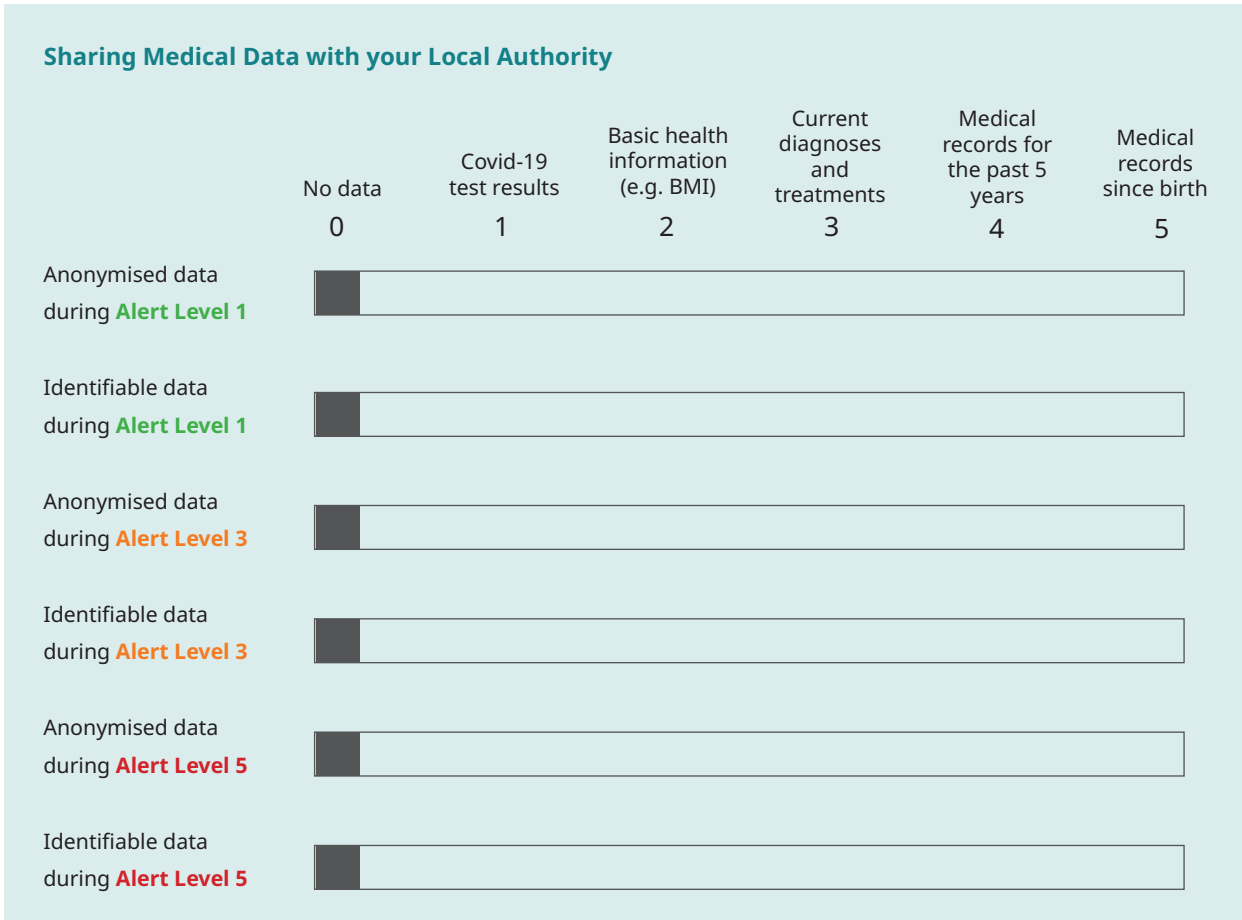


Figure 3.* Example of a Medical Data, Data Detail Slider question.

*The visual design of the question used in the study differs to the one shown here. Please refer to our WP3 snapshot report to view the actual design (<https://www.omddac.org.uk/wp-content/uploads/2021/08/WP3-Snapshot.pdf>).

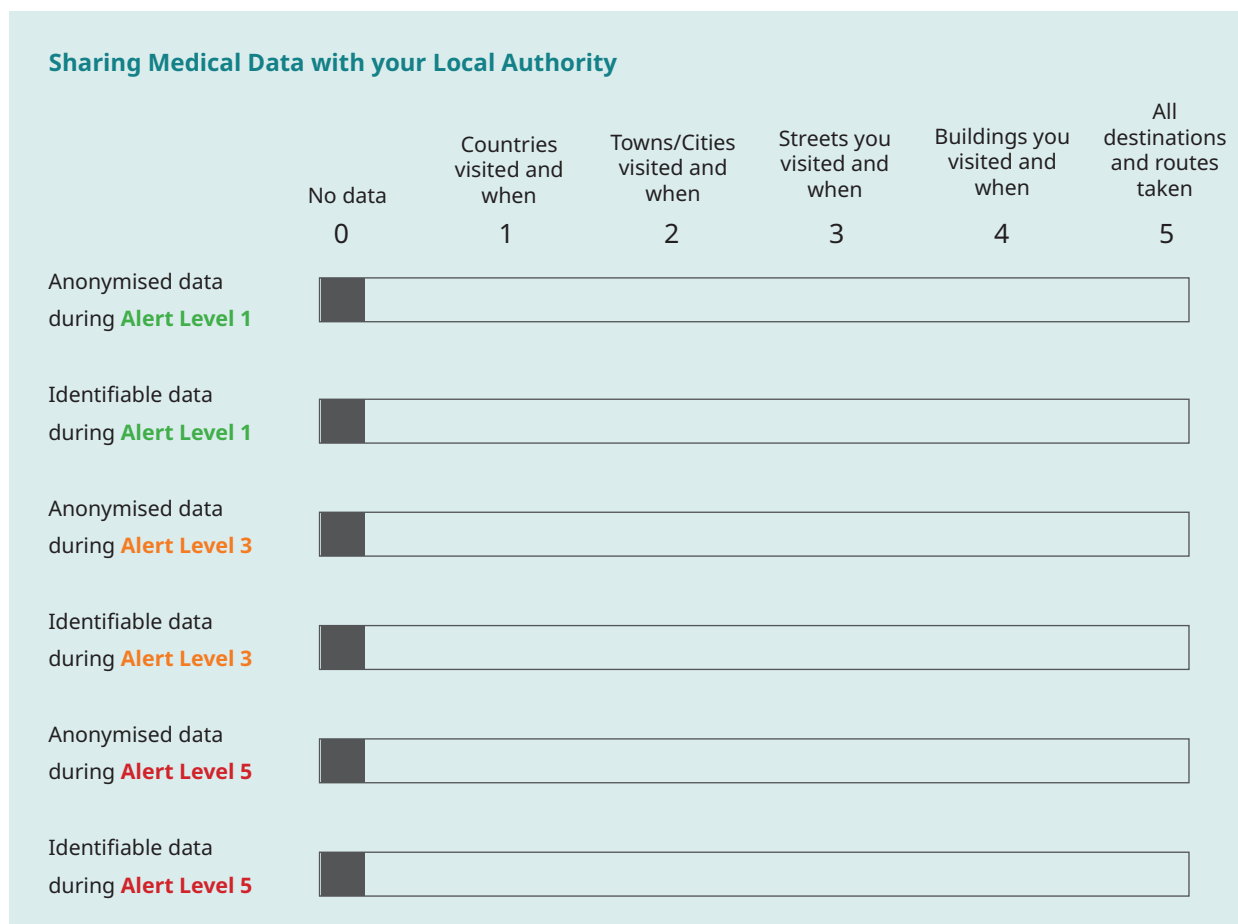


Figure 4.* Example of a Mobility Data, Data Detail Slider question.

Agenda Day and Online Questionnaire: Incorporating the Perspectives of Young People

To incorporate the perspectives of children and young people, OMDDAC worked with Investing in Children ('IIC'), a children's human rights organisation which works in partnership with children and young people to exercise their rights and participate in decisions that affect them. OMDDAC and IIC provided young people with an opportunity to express their views in two ways, by means of an agenda day or via an online questionnaire. An agenda day is an adult free meeting arranged by an IIC project worker and young facilitators to enable children and young people to meet and give their views on a particular issue. The agenda day took place virtually via Zoom on 6 July 2021. The online questionnaire was open from 23 June until 16 July, advertised to young people already in contact with IIC and via their Twitter account. A series of questions and scenarios were drafted by OMDDAC and provided to IIC for discussion with the young people, asking what young people knew about how the Government is trying to tackle COVID-19, and in particular what they knew about the way information and technology was being used to tackle COVID-19. They were asked for their views on the use of wastewater testing, inter-agency information sharing about individuals who are self-isolating, and the use of algorithms to determine examination grades. In total, 17 young people aged between eleven and eighteen provided their views, either at the agenda day or via the survey. Whilst this is a small-scale study and findings cannot be considered in any way representative of young people's views across England, it provides an important insight into the opinions and concerns of some young people.

*The visual design of the question used in the study differs to the one shown here. Please refer to our WP3 snapshot report to view the actual design (<https://www.omddac.org.uk/wp-content/uploads/2021/08/WP3-Snapshot.pdf>).



