



Social Mobility  
Commission



# **Data for social mobility: improving the collection and availability of data across government**

A review of data gaps and solutions to support effective policy-making

**Research report**

December 2022



© Social Mobility Commission 2022

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit [nationalarchives.gov.uk/doc/open-government-licence/version/3](https://nationalarchives.gov.uk/doc/open-government-licence/version/3)

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned. This publication is available at <https://www.gov.uk/government/organisations/social-mobility-commission>

Any enquiries regarding this publication should be sent to the Social Mobility Commission, Sanctuary Buildings, Great Smith Street, London SW1P 3BT.

Email: [contact@socialmobilitycommission.gov.uk](mailto:contact@socialmobilitycommission.gov.uk)

Front cover photo credit: a close up of two laptops on a table with some paperwork nearby and people working. Scott Graham at UnSplash.

# About the Commission

---

The Social Mobility Commission is an independent advisory non-departmental public body established under the Life Chances Act 2010 as modified by the Welfare Reform and Work Act 2016. It has a duty to assess progress in improving social mobility in the UK and to promote social mobility in England. The Commission board comprises:

## Chair

Katharine Birbalsingh CBE, Founder and Headmistress of Michaela Community School.

## Deputy Chair

Alun Francis OBE, Principal and Chief Executive of Oldham College.

## Commissioners

Dr Raghbir Ali, Senior Clinical Research Associate at the MRC Epidemiology Unit at the University of Cambridge.

Matthew Goodwin, Professor of Political Science at the University of Kent.

Ryan Henson, Chief Executive Officer at the Coalition for Global Prosperity.

Parminder Kohli, Senior Vice President EMEA at Shell Lubricants.

Resham Kotecha, Head of Policy & Government Affairs EMEA at Wise.

Rob Wilson, Chairman at WheelPower – British Wheelchair Sport.

## **Acknowledgements**

Main analysis and report writing by: Jude Hillary (Co-Head of UK Policy and Practice), Elizabeth Davies and Sharon O'Donnell (Associates) from the National Foundation for Educational Research (NFER) which is a leading provider of independent educational evidence and assessments in the UK.

Contributions by: Michael Pandazis and Simon Dilly from the Social Mobility Commission. This includes some findings and recommendations based on the Social Mobility Commission's own perspectives.

We are also grateful for contributions from: Nicola Di Luzio, Jo Johnston, Jill Sherman, Shannon Miller.

# Contents

---

About the Commission	3
Foreword	6
Executive summary	8
Introduction	17
Methodology	20
1. A vision for improving socio-economic data in the future	22
2. Data sharing – unlocking the power of administrative data	28
3. A single platform to access administrative data	33
4. Development of a household data set	36
5. Developing better measures of socio-economic background	40
6. Filling important data gaps	45
7. Process and practicality issues	53
Conclusion	56
Annex 1: Case study – a household data set to measure social mobility	57
Annex 2: Case study – proxies for disadvantage and social mobility	63
Glossary	67

# Foreword

---



Social mobility is a complex phenomenon. In order to improve it, we need to better understand it. Only by looking at good-quality data can we improve our interventions, and people's future outcomes.

The ability to collect and analyse socio-economic data is important to creating effective policy. Data allows us to identify the people and places which need the most help, and understand exactly what that help should be. Without data, we cannot make progress. But there are still 3 significant data challenges facing us.

First, data sharing, both across government departments and with external researchers, is often difficult or even impossible. Data that is held in Whitehall often cannot be accessed by others who could make use of it. Improving data-sharing capabilities and processes would allow many more people to use the data that already exists.

Second, data linking. The data we collect relates mainly to individuals. We need to connect the information we have on people living in the same household, and also connect one generation to the next. Linking data in this way would show us the impact of family circumstances on a person's socio-economic status, helping us to better understand the causes of social mobility, and target our interventions accordingly.

Finally, missing data. The UK still does not have administrative data on people's occupations. UK-wide birth-cohort studies, which have been so vital to social mobility analysis, have been infrequent and irregular, with just 4 carried out since the World War 2.

The government has made good progress in some areas. For example, the creation of the Inclusive Data Taskforce, Equality Data Programme and publication of the National Data Strategy are all welcome and have played a vital role in directing activity in the right areas.

The Pupil Parent Matched Database being developed by the Department for Education, which links pupil data to information about their parents or guardians, is also a good start. However, we would ultimately like to see a UK-wide household data set, covering every household. Only then can the government and researchers get a true picture of people's – and, critically, children's – circumstances.

This would not only allow policymakers to track households from a wider range of socio-economic backgrounds in a joined-up way, and to better target support and resources, but would also help to identify around 1.7 million 'lost children' living in relative poverty who are currently invisible in our education statistics. These children aren't eligible for free school meals, and there is no other currently available identifier of disadvantage, something that would be fixed by the ability to link their pupil record to their household income data.

As this report shows, there is still a lot of work to do to reach a future in which better socio-economic data can be easily accessed through a single platform, a UK-level household data set is in place, and socio-economic data is being fully used by policymakers and beyond.

However, with government action in a few important areas, and the sustained focus of the Social Mobility Commission, there is great potential for improving our data, and eventually, people's outcomes. While the findings and recommendations of this report may seem complex and technical, transforming our data has the power to transform people's lives. That's why we'll be reporting on government progress each year against the recommendations of this report.

# Executive summary

---

Improving social mobility and increasing opportunity for all has been a stated aim of governments for decades. We need a greater understanding of the causes of poor life outcomes to develop better-targeted policies. Data has the power to provide these insights, but often there are gaps in the existing data, barriers to cross-organisational sharing and with external researchers, issues with data linkage, and a lack of vital data sets. This limits the scope for potential analyses to shed light on the causes of socio-economic disadvantage and its implications on social mobility.

Policymakers and researchers in social mobility currently face really serious data challenges. For example, vital work on occupational mobility has principally relied on just 4 UK-wide birth-cohort studies. This is because administrative data on people's occupations has not been available, and the Labour Force Survey (LFS) has only recently begun to ask social-mobility questions. This means that calculating occupational mobility at a local authority level is very difficult. On income mobility, the situation is even worse. The LFS option does not exist for parental income, so there is no official data source for the rigorous analysis of income mobility. And on a related note, we don't have administrative data on the economic circumstances of children and the households they live in, so analysis and policy-making is limited to using eligibility for free school meals (FSM).

Even when data does exist, getting hold of it can be difficult. Sharing administrative data between public authorities is often a lengthy process. Often, one part of government isn't aware of what data is held by another part. External researchers still face practical barriers to getting safe access to data, leaving the UK at risk of falling behind in this vital area. Moreover, there are huge unexploited gains from linking data. While some progress has been made on these issues, there is still a long way to go.

The Social Mobility Commission's (SMC) vision is to significantly improve social mobility in the UK by helping the UK government understand how socio-economic circumstances are linked to life outcomes. The SMC commissioned the National Foundation for Educational Research to identify the current barriers to finding and using socio-economic data, to outline an ambition of what an improved data environment could look like and how it might



benefit future policy-making.<sup>1</sup> This research acts as an aspirational paper to stimulate discussion and engagement across the government. We provide a high-level overview of some of the challenges in the sharing and use of socio-economic data and highlight some of the benefits if these problems were addressed.

## **Better data sharing and linking**

### **Findings**

Sharing administrative data between public authorities can be a lengthy process, but there are tools and opportunities to improve this. All departments should be encouraged to share data through the Office for National Statistics' (ONS) Integrated Data Service (IDS), and access should be made easier for academics and other researchers to maximise the utility of these data sources.

The Digital Economy Act 2017 (DEA) is beginning to improve this situation, but sharing administrative data between public authorities can be a resource-intensive process. As part of fulfilling their responsibilities, government departments collect rich data which is important for understanding socio-economic differences. But sharing data between public authorities can be a complex business which takes significant time and resources to establish. Some existing data shares are bespoke in nature, limiting what the data can be used for, even though it could have benefits elsewhere. The relatively recent DEA has the potential to allow more and easier data sharing between public authorities. However, health and social care data, which would be valuable in understanding socio-economic disadvantage, is currently out of the scope of the DEA.

The IDS should help to join up cross-government administrative data on a single platform, but to maximise its value, all departments need to make their data available. During the COVID-19 pandemic, vital information could not be accessed as quickly as it was needed because the data required was held in separate data sets across the government estate. This adversely affected the ability to make quick decisions. The pandemic reinforced the need to be able to bring administrative data from across government together into one place or platform so that different data sets can be rapidly linked, as and when needed for specific purposes. The ONS has developed a new platform called the IDS to help address

---

<sup>1</sup> [National Foundation for Educational Research](#), 2022.

this situation. However, for this service to work effectively and to maximise the benefits to its users, departments and other important public authorities must make their rich administrative data available through this service. Also the ONS needs to make the process for users seeking access to data sets on the IDS as straightforward as possible.

Data sharing outside the government should also be improved in the UK, so that the quality of data available to researchers and the general public, and its accessibility, is improved. The experience of some non-governmental researchers suggests that the UK is not at the forefront of data sharing outside the government. Problems range from reluctance to share data, to practical barriers such as websites working slowly, or not at all, or having to re-apply for access every time a data set is updated. This damages the quality of research information available to the UK public.

### **Recommendations**

1. The ONS should promote the DEA information-sharing provisions across the government, so that data can be shared across departmental boundaries more easily and quickly. The ONS should also promote these data-sharing powers to academics and third-party researchers. Also, the possibility of including health and social care in the DEA should be explored to enable more joined-up analysis of socio-economic disadvantage.
2. Government departments should be strongly encouraged to make their administrative data available through the IDS so it could be linked to other administrative data sets. They should also ensure sufficient resources are made available to make this happen as quickly, accurately and securely as possible. The IDS should swiftly work towards DEA accreditation to support the linking and sharing of data by public authorities.

## **Filling the data gaps**

### **Findings**

Due to Universal Credit (UC) transitional arrangements, it is no longer possible to identify the length of time a pupil has been disadvantaged. Ordinarily, a pupil who is eligible for FSM would not remain so if their family income improved and they no longer met the eligibility criteria. However, to smooth the rolling out of UC, the government introduced

some transitional arrangements. Under these, any pupils who are eligible for FSM at any point between April 2018 until the end of the UC roll-out (which is set to be summer 2023 at the earliest) will retain their FSM eligibility for this whole period and until their phase (primary or secondary) of education ends. This applies even if their family circumstances improve and they would ordinarily no longer have been eligible for FSM. In general, the longer a pupil has been disadvantaged, the lower their average attainment, so it is important to be able to identify how long a pupil has been disadvantaged.

The main data gap identified to help better understand socio-economic differences was the need for a national household-level data set based on administrative data. Administrative data sources are increasingly being used by academics and researchers to explore socio-economic differences as these can allow for a more granular analysis of smaller geographical units. Much of the administrative data collected across government is about individuals, but to better explore socio-economic diversity, data is needed about a person's wider household. However, due to challenges with both data sharing and linking, there is no established administrative household data set. This limits the ability to understand the impact that policies have on all household members. Having a household-level data set would enable the government to look at an individual's or household's needs and target appropriate support and financial assistance accordingly, rather than the current unsystematic approach, which may be imprecise and poorly targeted, and could potentially result in different interventions undermining other initiatives.

One project which has the potential to overcome some of these challenges is the Pupil Parent Matched Data (PPMD), which is being developed by the Department for Education (DfE). This links pupil data to information about their parents or guardians. With the addition of income and geographic data, this has the potential to start to fill this gap in the future.

Existing survey evidence, such as birth cohort studies, the Labour Force Survey (LFS), and the Wealth and Assets Survey (WAS), has been critical for research and should be strengthened even further to collect more social mobility data. Although surveys will always have limitations related to sample size, they can provide rich, good-quality data to address questions that administrative data cannot, such as about the family environment. The last major British birth-cohort study was established in 2000, leaving a gap of more

than 20 years even if a new study could be set up immediately. We welcome UK Research and Innovation's (UKRI) endorsement of a new UK-wide birth cohort study.

The social-mobility questions on the LFS have proved to be very valuable for analysis and will be used for future annual reports by the Social Mobility Commission (SMC), while the WAS will provide important data to inform new measures of wealth mobility, but there may be room for strengthening these, for example, by adding questions on parental education or social capital to the LFS.

Improved data could lead to the development of better measures of socio-economic disadvantage which can improve the targeting of support. The eligibility for FSM and FSM eligibility in the last 6 years (FSM6) are well-established proxy measures for socio-economic disadvantage. However, research suggests these do not capture all of the households who need support or may be experiencing disadvantages.<sup>2</sup> Improvements in data linking and the development of household data sets such as the PPMD could make it possible to create new, more accurate measures of socio-economic disadvantage in the future. These new measures could be shared across government, enabling policymakers to track households from a wider range of socio-economic backgrounds in a joined-up way, and to better target support and resources.

Some important socio-economic factors such as data about a person's occupation are rarely collected or maintained by any government administrative system. To unlock the full power of data analysis to make inferences on the current and future picture of socio-economic disadvantage and its implications on social mobility, we need to be able to identify important areas for concern and have suitable metrics to track policy progress against. However, some vital aspects of socio-economic data are not routinely collected and maintained by any of the administrative systems across the government estate. For example, occupation is a key indicator of socio-economic status, which can be used to measure the social class of an individual. This is only collected in the UK population census every 10 years. Although, at the time of writing, HM Revenue and Customs

---

<sup>2</sup> UCL, Social Research Institute, '[Measuring socio-economic background using administrative data. What is the best proxy available?](#)', 2020. Published on ECONPAPERS.REPEC.ORG.

(HMRC) is consulting about whether to collect this data in future.<sup>3</sup> Other data gaps identified include parental income and education, hours worked by parents, non-earned income (for example, wealth), childcare arrangements, changes in household circumstances, and a child's non-cognitive skill levels. While these and other data gaps exist, this is likely to hinder improving our understanding of socio-economic inequalities. We welcome the recommendations of the Inclusive Data Taskforce to include more of this type of data in administrative data sets.<sup>4</sup>

Important research into income mobility between generations by geographic area is hampered in the UK by our inability to link parents to their children in administrative data sources. In many jurisdictions, the tax records of parents and children are linked. This has enabled detailed and high-quality analysis of income mobility, such as that carried out in the US by Chetty and others.<sup>5</sup> Unfortunately, there is no equivalent linking in the UK. Even worse, the second-best option of using the LFS is not available (as it is for occupation), because the LFS does not ask about parental income. This means that there is no official data source for the rigorous analysis of income mobility. We have to rely instead on academic panel studies, which suffer from major limitations (irregularity in the birth cohorts chosen, along with high attrition and lack of representativeness).

The government should therefore consider how such family linking might be feasible in the UK, to ensure that the UK does not fall behind other countries in its ability to analyse income mobility trends.

There is a lack of harmonised data available across the UK countries which affects comparisons. For some important measures, such as educational attainment and FSM eligibility, there is a lack of consistently comparable data across the UK nations. This is due to variations in policy and practice approaches, and to differences in the methodology used to calculate similar statistics and measures, which are used across the different UK

---

<sup>3</sup> Parental occupation is recorded on a child's birth certificate. Marriage certificates record details of the couple's occupations while death certificates collect information about the person who it relates to. However, all of these relate to a point in time and are not updated when a person's occupation changes.

<sup>4</sup> UK Statistics Authority, '[Inclusive Data Taskforce recommendations report: leaving no one behind – how can we be more inclusive in our data?](#)', 2021. Published on UKSA.STATISTICSAUTHORITY.GOV.UK.

<sup>5</sup> Raj Chetty and others, '[The fading American dream: trends in absolute income mobility since 1940](#)', 2017. Published on PUBMED.GOV.

nations. A lack of harmonised data hinders making informed comparisons across the UK to help support policy-making in both the devolved nations and across the wider UK government.

## **Recommendations**

3. A UK household data set based on administrative data would be very desirable. The government should set up a central programme to work towards developing such a data set. This should make use of existing data developments like the PPMD, if possible, to get something which can be used across government and beyond immediately. The programme should have an agreed, prioritised plan for further development to serve a wider set of user needs and move us towards our ambition to have a full UK household data set based on administrative data. Critically, a household data set that included children would allow much more accurate targeting of education policy than the current deprivation measure, based on FSM eligibility.

4. It is important to collect information to identify the length of time a pupil has been disadvantaged as attainment outcomes tend to be lower on average for the more persistently disadvantaged. The DfE should undertake a review to identify how best to collect this information, bearing in mind it may already be held by or can be derived from information held in the Department for Work and Pensions (DWP).

5. The ONS, with SMC input, should lead a cross-government project to conduct research into developing improved social background measures which can help form a wider picture of the full range of socio-economic characteristics experienced across the population. These should be developed for the use of both policy and operational purposes. These could be shared across government, enabling policymakers in all departments to track households from a wider range of socio-economic backgrounds in a joined-up way, and to better target support.

6. The government should implement current proposals to collect and – critically – share occupational data which is needed for a better understanding of social mobility and the causes of socio-economic disadvantage. This could be collected by HMRC from employers through the pay-as-you-earn and national insurance system. If occupational data is not collected, any household data set would be lacking critical information.

7. Linked administrative income and education data between parents and children is an important evidence gap for social-mobility analysis. Without this, researchers cannot look at the earnings or education of today's adults and compare them with the earnings or education of their parents. There is currently no clear way forward to achieving this in the UK. The National Statistician should work with the SMC and other departments to identify and assess potential solutions for developing administrative data systems, and to put in place a programme of work to address this important evidence gap.

8. We welcome UKRI's endorsement of a new UK-wide birth cohort study as a successor to the 4 studies that have taken place since 1945. The government should consider how this critical work can be best supported now and in the future, so that further large gaps do not arise.

9. The government should commission the National Statistician to lead a cross-government review of UK nations and in partnership with the devolved administrations. This would identify the important data gaps that are hindering our understanding of social mobility and how they could be addressed.<sup>6</sup> This review should take both administrative data and survey data (such as the LFS and the WAS) into consideration.

## **Leadership and collaboration**

### **Findings**

Policy and operational responsibility for addressing social mobility and socio-economic disadvantage are dispersed across the government and not always joined up. No single, specific part of government is currently responsible for overseeing all of the work being undertaken to tackle social mobility. Each area of government is working to achieve their own specific objectives, but greater progress might be achieved if there was greater oversight and greater collective work to improve social mobility outcomes. According to PricewaterhouseCoopers, if central accountability for cross-government social mobility was strengthened, it could mitigate the risk of unintended negative consequences from

---

<sup>6</sup> The National Statistician the Chief Executive of the UK Statistics Authority, Permanent Secretary of the Office for National Statistics and Head of the Government Statistical Service.

individual policy plans.<sup>7</sup> Given the current cost of living crisis and the government's levelling up policy, which seeks to reduce the imbalances between areas and social groups across the country, this is a good time to review how this important plan is overseen and managed.

### **Recommendation**

**10.** As an arms-length body, the SMC does not have the power to require government act on the recommendations it makes. Therefore, we encourage the government to assign the recommendations in this report to an appropriate central body that can take coordinated action on them. For example, the Central Digital and Data Office responsible for the National Data Strategy or the ONS Centre for Equalities and Inclusion.

In addition to the previous recommendations, the appointed lead should also be responsible for publishing a regular audit of the non-sensitive data sets held in each department (although the release of the actual data, where appropriate should be done in a properly planned way, and not necessarily as part of this audit).

To help deliver on these recommendations, the government should also give a relevant department such as the Cabinet Office or the Department for Levelling Up, Housing and Communities' overall responsibility for oversight of work across government to address socio-economic disadvantage. This would encourage departments to set meaningful objectives on social mobility data gaps and social mobility more generally, and hold departments to account for delivery.

---

<sup>7</sup> PricewaterhouseCoopers, '[How government can drive social mobility in the UK](#)'. Published on PWC.CO.UK.



# Introduction

---

This report has been prepared by the National Foundation for Educational Research on behalf of the Social Mobility Commission (SMC). It provides findings and recommendations from desk-based and qualitative research to explore what the priority areas for improving socio-economic data across government should be and to identify solutions for addressing these areas.

There are 3 key challenges the government faces in designing and implementing policy to help those from the poorest socio-economic backgrounds and to improve social mobility. First, it can be difficult to identify the groups of people who require the most help. Second, it can be challenging to determine the optimal approach to design and target a policy to help those who need it most due to a limited evidence base. Finally, once a policy has been implemented, measuring progress and identifying whether it is working is also difficult.

Although there is no simple solution to these challenges, they do have one common underlying problem: the availability of appropriate socio-economic data. This report acknowledges there are many other factors which provide a challenge to the government, such as the political context in which policy-making occurs and the difficulties associated with the financial and resource constraints within which the government must operate. However, it is clear that an improvement in the availability, quality and type of socio-economic data available, has the potential to significantly ease some of these challenges in policy-making and improve the targeting of support and resources. This report sets out an aspiration for what improvements in the availability of socio-economic data could look like. It does not provide prescriptive solutions on how to achieve this vision, nor does it provide rigorous detail on the data challenges faced or the costs versus benefits of delivering different parts of this vision. The aim of this report is instead to shed light on the issue of socio-economic data in government and stimulate debate and action on addressing its availability to help social mobility policy-making in the future.

## Project aims and objectives

Ensuring the government has gold-standard data to understand inequalities and the distributional impacts of policies by socio-economic background is a key milestone to improving policy-making in this area. The SMC is undertaking an ambitious and innovative data programme to reform the availability, quality and use of socio-economic data across government.

The purpose of this report is to highlight some of the challenges with using socio-economic data to inform policy, to discuss how to overcome these, and to highlight the benefits for social mobility. It also seeks to identify how to improve the data to better identify subgroups who face specific challenges, or who have achieved better social mobility outcomes. We want to engage key decision-makers and leaders in data across government in conversation and action on improving the access, quality and linkages of socio-economic data.

However, the report is not intended to provide a comprehensive view of the data landscape. We acknowledge that there may be other projects within government and other issues which this report has not covered which may be of value to improving our understanding of socio-economic disadvantage. We encourage any teams working on such projects to contact the SMC.<sup>8</sup> This report is not intended to provide precise solutions on how to overcome these challenges, as these technical details would be best tackled in a future, more detailed review.

The aims of this research is to set out an aspiration for how socio-economic data in government could be improved. For example by: setting out the current state of socio-economic data availability; providing a vision for the socio-economic data environment we could aspire to have; highlighting some key areas through which socio-economic data could be improved; and providing high-level recommendations to suggest how these improvements could be made.

---

<sup>8</sup> [contact@socialmobilitycommission.gov.uk](mailto:contact@socialmobilitycommission.gov.uk).

There are multiple government departments and other public bodies which have an interest in socio-economic data. However, the scope of this research project was restricted to some select departments and policy areas. These include: the Department for Education – as education can break the link between a child’s social background and their later outcomes, improving social mobility; the Department for Work and Pensions – whose focus is on welfare support, both in and outside of work, and on tackling child poverty; and the Department for Levelling Up, Housing and Communities – who lead on geography and place measures, which are important in understanding social mobility.<sup>9</sup>

In addition, we consulted with a senior official at the Office for National Statistics to discuss some of the potential solutions to challenges faced, particularly around data sharing and the role the Digital Economy Act 2017 could play in addressing this, and about the development of its new Integrated Data Service.

---

<sup>9</sup> Please note that this department was called the Ministry of Housing, Communities and Local Government when this project started, but it has since changed its name. We refer to it by its new name in this report, the Department for Levelling Up, Housing and Communities.

# Methodology

---

The research involved the following key elements:

A rapid evidence review including relevant ministerial statements, policy documents, research evidence and other documentation relating to the agreed departments and policy areas. These were used to assess which data sources are available for monitoring social mobility and to understand the causes of a lack of social mobility in England. They also provided insights into the existing literature on the data limitations and important unanswered questions surrounding social mobility and socio-economic inequalities.

Consultation with departmental policy leads to identify views about what policy questions currently cannot be fully answered with existing data, and what data improvements are needed to better look at the socio-economic impacts of policies.

Consultation with departmental analysts to discuss what the current data limitations are that impede understanding of socio-economic inequalities and how they could develop data sources to address these barriers.

A total of 28 interviews were completed with officials across the respective departments. Findings from the different interviews were drawn together into a gap analysis framework to inform the development of this report (see accompanying annexes).

## Report structure

This report explores the current capability of the system for capturing and processing socio-economic data, key areas for change, where improvements have been made or are in progress, and where further improvements may be beneficial for understanding social mobility and socio-economic disadvantage.

The remainder of the report is structured as follows:

**Chapter 1:** Provides a vision for improving accessible data through investment in better data systems to strengthen the government's understanding of socio-economic circumstances.

**Chapter 2:** Explores the challenges of data sharing and how administrative data might be unlocked for wider use.

**Chapter 3:** Examines the need for a single platform for data sharing and analysis, and the opportunities presented by the Office for National Statistics' Integrated Data Service.

**Chapter 4:** Looks at the need for opportunities to link data about an individual to other members of their households to improve insights.

**Chapter 5:** Examines the current use of proxies for disadvantage and how these may be improved.

**Chapter 6:** Looks at where the collection of additional data items might be beneficial in our understanding of socio-economic disadvantage.

**Chapter 7:** Explores some wider practical issues that may limit the analytical process.

**Annexes 1 and 2** provides detailed case studies exploring issues identified through our interviews which were of particular importance to respondents. These were the challenges relating to linking data sources together to form a household-level data set, and establishing better, more accurate alternative measures for socio-economic disadvantage. These 2 interrelated issues, the possible solutions, and the benefits to improving these areas are explored in more detail.

# 1. A vision for improving socio-economic data in the future

---

## Why socio-economic data matters

It has been a stated aim of successive governments to increase social mobility and tackle child poverty. Ideally, socio-economic data would not matter, because no group would face greater barriers than any other in fulfilling their potential. However, as this is still not the case, even in advanced economies, it is important to maintain a lens on living conditions, wages, and opportunities. It matters because low social mobility and widespread socio-economic disadvantage can undermine economic growth as the most disadvantaged young people do not reach their full potential. A lack of mobility can affect people's physical and mental health, life expectancy and life satisfaction, which potentially could affect social cohesion.<sup>10</sup>

There has been much investment of effort and resources by governments for decades, which has resulted in some small improvements in narrowing disadvantage gaps. However, any improvements achieved to date are likely to have been reversed by the pandemic, which has disproportionately affected the most disadvantaged in society. For example, pupils from deprived backgrounds suffering greater disrupted learning than their more advantaged peers.<sup>11</sup> As highlighted in the Levelling Up White Paper, there is significant geographical inequality in the UK, which is one of the most spatially unequal countries in the Organisation for Economic Co-operation and Development.<sup>12</sup> Looking forward, there are also concerns that the Fourth Industrial Revolution will cause disruption to labour markets in the future, and will likely compound differences in socio-economic inequalities in countries unprepared to take advantage of new opportunities.<sup>13</sup> There is

---

<sup>10</sup> The Organisation for Economic Co-operation and Development, '[A broken elevator? How to promote social mobility](#)', 2018. Published on OECD.ORG.

<sup>11</sup> National Foundation for Educational Research, '[COVID-19 recovery](#)'. Published on NFER.AC.UK.

<sup>12</sup> HM Government, '[Levelling Up the United Kingdom](#)', Published on GOV.UK.

<sup>13</sup> World Economic Forum, '[Global Social Mobility Index 2020: why economies benefit from fixing inequality](#)', 2020. Published on WEFORUM.ORG.

now as strong a case as ever to maintain focus on tackling disadvantage to make a lasting difference.

The Social Mobility Commission's (SMC) new Social Mobility Index captures a range of evidence-based drivers of social mobility and both intermediate and longer-term outcomes, to track progress and inform policy. Socio-economic data is a crucial input to the index because it helps identify areas and population subgroups which may be at risk of experiencing poorer outcomes. Without good data to identify the key areas of concern, it is difficult to build an evidence base on which policies may work, as measuring progress in absence of the requisite data is challenging. This leads to a vicious cycle in which the people requiring support are not identified, the evidence around which policies work is not developed, and effective policy solutions may not be implemented.

### **Inequality and social mobility**

**Inequality measures** look at how socio-economic factors, such as income, wealth or education, are spread out across the population. For example, what is the income of high earners versus the income of low earners? Or how is education spread among the population?

**Social mobility measures** may look at similar factors, but they consider how they change over an individual's lifetime and the change between an individual and their parents. For example, how many people earn more than their parents did at a similar age? Or how many people are in a different occupational class from their parents?

The availability and access to appropriate socio-economic data has the potential to significantly improve the social mobility landscape through multiple channels. First, socio-economic data is often the first step in evidence-based policy-making, as data analysis forms a crucial part in the production of robust evidence. Simply put, without the appropriate data, it is difficult to develop the evidence which in turn makes it even more challenging to design the policy aimed at a specific problem. Unlocking the full power of socio-economic data may allow for the identification of the people or places that require the most help. Analysing socio-economic data may also help us find people or households in a population subgroup or part of the UK who are 'bucking the trend' and achieving outcomes which are better than would be predicted for someone in their socio-economic

situation. This could in turn help identify future potential policies which could be targeted to help bring opportunities to people of similar socio-economic backgrounds in other parts of the UK.

## **The current state of play**

Currently there are 2 key issues:

### **Longitudinal survey data**

Historically, much of the research work done to study intergenerational mobility has used longitudinal studies such as the National Child Development Study (which began in 1958) the British Cohort Study (1970), or the Millennium Cohort Study (2000), which follow a sample of people at regular intervals. These important studies contain rich data that have helped to identify some of the key drivers of social mobility. They provide information that administrative data sets cannot, and they also make intergenerational comparisons easier than most administrative data. Information is collected from an early stage on children, their parents, and sometimes grandparents too. Some of the best studies about intergenerational mobility have come from such data sets. They can also be linked to administrative data.

However, a limitation of survey data is that even if the original sample size is seemingly large, once the sample is split by geography and protected characteristics, it can be too small to use to make inferences. This limits the ability to use large survey data for more detailed analysis of population subgroups and regions and analyse the intersections with socio-economic factors. The birth cohorts in these studies are also infrequent and irregularly spaced – the last one was in 2001 – and they suffer from high attrition and problems with representativeness. This creates serious data gaps. Having more regular birth-cohort studies would go some way to addressing this.

As the power of information technology has grown, greater focus has been placed on utilising the rich administrative data that government departments collect as part of their functions to understand socio-economic differences in society. However, data sharing across departmental boundaries is often difficult because of the complexity and time needed to put the necessary legal powers in place. Sometimes, insufficient time is set aside for setting up new data-sharing arrangements and there can be a lack of shared



aims between the parties involved. But the process is very helpful to identify people who most need public services, so this is definitely worth the investment.

## **Lack of household data**

The second issue is that many of the key administrative data sources collect data about individuals rather than their households. This means that creating household-level data can be difficult as many administrative data sets do not collect the data needed to make these linkages. So it can be challenging to measure household income or wider family circumstances, which are key pieces of information for understanding the socio-economic circumstances of a household.<sup>14</sup>

In their absence, measures such as free school meals (FSM) eligibility are used to identify the disadvantaged families to support. However, these measures are far from perfect as they do not identify all of the families who would benefit from support. FSM is also a binary measure, dividing families into two groups so that all nuance is lost. There are also several important gaps in the administrative data, such as occupation and hours worked, which are not collected routinely by the government. These factors combine to hinder efforts to better understand drivers of low social mobility and socio-economic disadvantage.

## **A future vision**

To enable the development of effective policies to improve social mobility and tackle disadvantages, policymakers and analysts first need to have access to key data which can help them to better understand the underlying causes. However, there are some areas where action could be taken to improve the current landscape, including developing data-sharing legislation and exploiting existing legislation, investing in the collection of new

---

<sup>14</sup> The concept of a household, as defined by the Office for National Statistics (“one person living alone or a group of people, not necessarily related, living at the same address who share cooking facilities and share a living room or sitting room or dining area”) is challenging to meet when using administrative data. This is because administrative data is nearly always collected at an individual (one row per person), or transaction level (for example, one row per benefit claim, one row per hospital visit), whereas survey collections can be designed specifically to collect information about households. Most administrative data research uses the concept of an ‘occupied address’, which groups individuals on administrative data who are recorded as living at the same address.

data, enhancing data linking and creating new data sets. We have created a vision to bring to life what the landscape could look like in the future. Through the following statements, we describe how different the future state could be if this vision was achieved:



**Legislation improvements have made sharing data easier:**

administrative data – including health and social care data – can be shared more easily across government and beyond, in a safe, secure and timely way. This allows analysts to develop greater research insights, which in turn helps policymakers to generate more effective policy solutions that are thoroughly evaluated. Operations are also better joined-up so that services are delivered efficiently.



**Cross-government administrative data can be more easily accessed through a single data platform:**

the development of the Office for National Statistics' (ONS) Integrated Data Service helps users within and outside of government to bring together administrative data sets held by different public authorities more easily. These data sets can be linked together more efficiently, accurately and safely using ONS's reference data management framework.



**There is consensus on data definitions and measures used across government and harmonisation of key data across UK nations:**

work carried out to agree and implement the use of common data definitions, key terms, measures and methods within government and beyond is helping users gain better insights into social mobility and socio-economic disadvantage. Better harmonised data across UK nations makes it easier for users to monitor progress across countries.



**A new administrative UK-level household data set provides greater insights:**

a UK-level household data set has been developed, which incorporates key linked data from a range of administrative sources held across government and beyond. An anonymised version of this data set is made available to departments, which helps develop more overarching, cross-government policies to support disadvantaged households.



**A more reliable set of measures of socio-economic status is in use across the government:**

using the UK household data set, improved social background measures have been explored and developed by a cross-government group for both research and operational purposes. Coalescing around one or more

common disadvantage measures, departments can identify and target the same families who need support and look at developing joined-up policy solutions and improving delivery.



**New administrative data is helping to enhance insights:** investment in the collection of new administrative data items which are known indicators of socio-economic position (for example, occupation) or could be important factors in the future are helping to better identify and understand the distribution and causes of socio-economic disadvantage.



**Socio-economic data is being fully exploited and used by policymakers across government:** civil servants have access to up-to-date, gold-standard data, which is available at the lowest possible spatial level and enabling them to better understand the causes and distribution of low social mobility and socio-economic inequalities. They can then develop effective policies to tackle this.



**Socio-economic data is being fully exploited to fill evidence gaps:** opening up linked administrative data sources and the UK household data set, and making them available to academics and third-party researchers, as well as government analysts, is helping to increase the potential resource available to exploit the data fully and produce new insights to fill gaps in the evidence base.

Taken together, these statements show there is great potential to achieve a step change in how we can develop data systems and processes to improve our understanding of the causes of socio-economic disadvantage and to develop solutions for addressing these. The remainder of this report reviews the current position in these areas and considers what more needs to be done.

## 2. Data sharing – unlocking the power of administrative data

---

Effective data sharing across government could help unlock the full potential of socio-economic data. This could significantly improve understanding of how socio-economic factors are distributed across areas in England and population subgroups. However, to achieve this, there need to be improvements in the sharing of data between government departments and other public bodies. Without effective data sharing, it is a challenge to consider the impacts of a policy in one area on another (for example, the impact of education on health). In some circumstances, important socio-economic data may only be held by one government department and without linking, this variable cannot be effectively included in policy thinking in other departments. To enable improved data-sharing practices across government, there are various hurdles which need to be addressed.

While the sharing and linking of data between departments and other public bodies provides the opportunity to increase the utilisation of the data and derive important new insights, there always needs to be a strong rationale for sharing data, underpinned by a proper legal basis. Also, to retain public confidence, all parties need to take great care to ensure the data they hold and have access to is kept securely and only used by accredited researchers who are trained to use the data safely and securely.

### **The barriers to data sharing**

There are legal barriers which can limit data sharing across government. Much of the administrative data which is useful for this plan is held by different departments of state. For example, the Department for Education (DfE) collects data about children and young people's education; the National Health Service (NHS) collects data about peoples' health and social care; HM Revenue and Customs (HMRC) collects earnings data and the Department for Work and Pensions (DWP) collects benefits data. But bringing data together owned by different departments is far from straightforward as public authorities need to have legal powers to share data. Sometimes it is necessary to seek changes to primary legislation to obtain the powers needed to share data between public authorities, and this can take years to bring about.

Data sharing agreements are sometimes limited to certain uses. Current data sharing agreements are often bespoke in nature, set up between 2 or more departments to address a specific set of research questions, to evaluate a particular policy, or other purpose. Unless wider data sharing is specified in the data-sharing agreement, this means that the linked data set which is created as an outcome of the data sharing cannot be used more widely than for the bespoke reasons it was created, even when it could be beneficial to other areas of government. The resource and length of time needed to put data-sharing agreements in place can sometimes leave less time and resources available for analysing the final linked data set, risking limiting the impact the new data set has on policy-making. There can also be challenges in opening up data sets to academics and third-party researchers who can help add to the analytical resource and expertise to exploit data sets fully. Therefore, these data sets are not always used to their full potential.

Perceived disadvantages of data sharing can sometimes reduce a data owner's appetite to support a request. For example, some ministers and senior officials in departments may be wary of data breaches or other risks to data security especially as a breach may lead to a fine or investigation from the Information Commissioner's Office. Some ministers may also be wary about public concerns about privacy in relation to 'big data' initiatives. As a result, there may be limited appetite to share data, despite the benefits it could bring about. As data-sharing agreements are often voluntary, some data owners may be reluctant to engage in sharing requests.

Even when departments want to share, there are barriers to linking the data. Many government departments maintain their own unique identification number, which they use for managing data. For example, the DfE uses its unique pupil number to identify pupils while they are in the school system, HMRC and the DWP use the National Insurance number, and the Department of Health and Social Care use the NHS number. Therefore, linking individuals across these administrative data sets is difficult as there is no single unique identifier in use across government. Analysts instead use a method called 'fuzzy matching' to overcome this challenge, but this approach is imperfect as it can result in some cases not being matched or matched incorrectly, which can affect some groups more than others.

Analysts in government departments may not always be aware of data collected by other departments which could be relevant to them. Some information about data held by departments is published on gov.uk, but it is not always up-to-date, accurate and complete. This means that existing data may not be being fully used. A centrally-held list would go some way to addressing this, but would need to be kept up to date to be useful.

## **Data-sharing developments**

There have been several data sharing and legislative developments since the last Social Mobility Commission report on data gaps in 2015, which identified data sharing as a key barrier.<sup>15</sup> The following 2 in particular are noteworthy.

Firstly, the Small Business, Enterprise and Employment Act (2015) enabled the linking of DfE and the Higher Education Statistics Agency's education attainment outcomes data with pay and benefits data from HMRC and the DWP respectively. This led to the creation of the new Longitudinal Education Outcomes (LEO) data set, which enabled analysts and researchers to track cohorts of pupils longitudinally from education into the labour market for the first time. Initially, access to LEO was strictly limited to government departments and a small number of external organisations. This situation has improved recently, and academics and third-party researchers can now apply to access some of the LEO data securely through the Office for National Statistics' (ONS) Secure Research Service (SRS).

The second substantial development has been the passing of the Digital Economy Act (DEA) 2017. This Act aims to improve public services through the better use of data, while ensuring privacy, clarity, and consistency in how the public sector shares data.<sup>16</sup> Part 5 of the DEA focuses on digital government, providing gateways that allow specified public authorities to share data with each other for research and operational purposes (see the DEA 2017 box for further information). Indeed, the recent improvement in access to the LEO data set through the SRS has been achieved under the DEA powers.

---

<sup>15</sup> Social Mobility and Child Poverty Commission, '[Data and public policy: trying to make social progress blindfolded](#)', 2015. Published on GOV.UK.

<sup>16</sup> Information Commissioner's Office, '[Data sharing across the public sector: the Digital Economy Act codes](#)', Published on ICO.ORG.UK.

**The Digital Economy Act 2017** received Royal Assent, becoming law, on 27 April 2017.

It makes provision for electronic communications, infrastructure and services. This covers a range of areas including access to digital services, digital infrastructure, intellectual property and digital government. Of particular note, Part 5 of the Act includes several chapters designed to improve the sharing of publicly held information for specific purposes, including:

**Public service delivery:** allows data sharing to support services and interventions for citizens and households as part of delivering social and economic policies. For example, supporting households with multiple disadvantages, people living in fuel or water poverty.

**Fraud and debt:** allows organisations to quickly establish data-sharing pilots to test the value of data sharing to combat fraud and tackle debt owed to the government.

**Data sharing:** permits public authorities to share de-identified information with accredited researchers for research in the public interest. However, this does not cover data sharing relating to the provision of health and social care.

The information-sharing provisions became fully operational in July 2018.

The DEA information-sharing provisions should be a big step forward in giving departments and external researchers the ability to share data more easily between themselves for research than in the past. After a slow start, partly due to the need to establish an operational process for making and assessing data applications and needing to train and accredit researchers in the safe use of data, the new system and processes have now bedded in and are making an important difference.

## **How can data sharing be improved further?**

The DEA has the potential to provide the legislative solution to enable better data sharing between departments, both for research and operational purposes, but there are still some issues which need to be addressed.

The DEA may promote greater data sharing between departments for operational purposes as well as for research needs. This may help with the delivery of better and more cost-efficient public services while also reducing data collection burdens.

The DEA does not currently cover data sharing relating to the provision of health and social care, which is a major limitation with regard to developing a rounded picture or understanding of socio-economic disadvantage. Steps should be taken to address this.

While the DEA has had a positive impact on timescales to share data, arrangements can nonetheless still take a reasonably long time to put in place.

Government researchers who want to access data for research purposes have to go through the same extensive checking and accreditation process as non-government researchers, despite being Crown employees doing government business.

It may be the case that some analysts and researchers, both inside and outside of government, are not yet fully aware of how the DEA can help them obtain access to cross-government data sets.

### **Recommendation 1**

The ONS should promote the DEA information-sharing provisions across the government, so that data can be shared across departmental boundaries more easily and quickly. The ONS should also promote these data-sharing powers to academics and third-party researchers. Also, the possibility of including health and social care in the DEA should be explored to enable more joined-up analysis of socio-economic disadvantage.



### 3. A single platform to access administrative data

---

The COVID-19 pandemic has shone a light on the need to bring data together from across the government, particularly health data, to better understand the impact of the virus on the country and assist in planning. As with the social mobility challenge, much of the data needed was held on separate data sets in different departments and vital information could not be accessed as quickly as needed, affecting the ability to make quick decisions. Due to necessity, some tactical, makeshift solutions were used to bring the necessary data together during the pandemic, but the crisis pointed to the need for a more strategic solution.

#### **Benefits and demand for a single point of access**

Discussions with policy leads and analysts during the interview phase of this study explored a range of issues and areas of interest for different departments, where further information would be helpful for their policy development and planning activities. This includes subjects such as the increased use of foodbanks, patterns associated with homelessness and rough sleeping, and understanding the impact of health episodes or interaction with the criminal justice system on school attendance. It also includes high-quality data about the background characteristics of groups of people who are receiving or being targeted for help, which may be improved by drawing together different administrative data held across the government. For example, using ethnicity data collected on the UK population census to supplement administrative data sets where this is patchy or only collected at the 5 broad ethnic categories (Asian, Black, Mixed, White, Other).

Departments would be able to better design and target support if they could understand the underlying contexts, characteristics and experiences of people engaging with these systems and services. It was felt that best practice approaches for access to, use, and linking of data should be determined and shared to support this understanding.

## The Integrated Data Service (IDS)

To meet these needs with a strategic solution, the Office for National Statistics (ONS) received funding to develop a new IDS which would bring together ready-to-use data sets from across the government and readily available non-government data that can be linked together for specific purposes. This has the potential to simplify and maximise access to a range of data while balancing legal, security, data protection and ethics, and wider obligations. The service enables accredited analysts and researchers from across government, as well as academia and research organisations, to seek and get access to multiple data sets which can be linked in the IDS and be made available for undertaking new research and producing new analytical outputs.

Central to the development of the IDS is the ONS's plan to develop a reference data management framework, which consists of a series of identifier variables for linking data sets for example, at the individual, household or employer level. This may enable data sets held on different departmental servers across the government to be linked together more efficiently as fewer resources are required to do the linking. The linked data sets may be more accurate and complete, compared to those created currently using the commonly used approach of 'fuzzy matching'. This is because 'fuzzy matching' involves matching individual records across data sets using a combination of variables other than identifier variables, such as the date of birth and address, to match individual records with some degree of confidence. A challenge with this method is that it can result in some data for some individuals not being matched, which can affect some subgroups of the population more than others. It is also possible that some cases may be incorrectly matched together. Another advantage of the IDS is that it should reduce data security risks as fewer personal data items (such as date of birth or address) would need to be shared to link the data sets.

The creation of the IDS has the promise to overcome significant hurdles which government departments face in accessing and linking their data to other departments, and to external data. But to maximise the benefits of this system, all government departments must commit to making their rich administrative data available through the IDS.

It is also important that chief analysts and heads of data in departments are aware of the ONS's IDS development programme. They should be proactively considering what opportunities this will generate in terms of improving their understanding of

socio-economic factors and planning what resources will be needed to do this. Chief analysts and heads of data should be making the case internally within their departments to ensure that their data is accessible through the IDS so it is available for linking to other departments' administrative data sets. This will help to produce new evidence and powerful insights.

### **Recommendation 2**

Government departments should be strongly encouraged to make their administrative data available through the IDS so it could be linked to other administrative data sets. They should also ensure sufficient resources are made available to make this happen as quickly, accurately and securely as possible. The IDS should swiftly work towards DEA accreditation to support the linking and sharing of data by public authorities.

## 4. Development of a household data set

---

A number of the key administrative data sources in government relate to an individual adult (for example, HM Revenue and Customs (HMRC) income data) or child, (for example, the Department for Education's (DfE) National Pupil Database).<sup>17</sup> However, information is not routinely collected in these administrative data sets about a person's wider household structure. But this is needed to understand the different factors that may contribute to social mobility outcomes.

The development of a household-level data set was identified as a key gap which needed to be addressed to enable departments to understand a person's household context better. This should enable departments to understand what impact their policies are having on all household members, which in turn could help to refine policies and improve the delivery and effectiveness of the support they provide.

### Development of the Pupil Parent Matched Data

There has been work done to develop household-level data sets based on administrative data sources before. For example, the Office for National Statistics (ONS) has undertaken some research to produce occupied address measures from administrative data.<sup>18</sup> This has been used in the development of experimental administrative data-based income statistics, which provide both individual and occupied address level measures of income.<sup>19</sup> The ONS has also published feasibility research on deriving highest level of qualifications from administrative data.<sup>20</sup>

---

<sup>17</sup> The National Pupil Database is a register data set of all pupils in state schools in England. It contains attainment data as children progress through school, as well as information on pupil background, absences and exclusions from school.

<sup>18</sup> Office for National Statistics, '[Occupied address \(household\) estimates: further information](#)', 2021. Published on ONS.GOV.UK.

<sup>19</sup> Office for National Statistics, '[Admin-based income, England and Wales: tax year ending 2016 revised results](#)', Published on ONS.GOV.UK.

<sup>20</sup> Office for National Statistics, '[Admin-based qualification statistics, feasibility research: England](#)', Published on ONS.GOV.UK.

The Supporting Families Programme, which is a cross-government initiative located in the Department for Levelling Up, Housing and Communities (DLUHC), created a new data set called the National Impact Study, comprising administrative data from the Department for Work and Pensions (DWP), the DfE, and the Police National Computer at the Ministry of Justice (MoJ). However, this data set was primarily created for evaluating this programme, so only contained records for families being supported by it along with a sample of similar families who were not involved in the programme. Therefore, the data set has less value for analysing other questions.

Another interesting initiative, which the DfE is taking forward, is its Pupil Parent Matched Data (PPMD) development. This links learner records from the national pupil database and individualised learner records to parents and guardians using Child Benefit claims initially and information from Tax Credits and Universal Credit claims where available. This is linked to earnings data from HMRC and benefits data from the DWP for the parents and guardians to gather information about household income. Bringing this data together starts to create a household-level data set which can be used to produce new socio-economic insights, such as potentially identifying more accurate measures of disadvantage.

The PPMD project is at a relatively early stage of development. While there is much development work planned and issues to address, it has the potential to become the type of household-level data set that departments say they need. The DfE has plans to extend the data set longitudinally by linking it to other data sets, such as the Longitudinal Education Outcomes data set, which could provide information about how parental income changes over the years that their children are in the education system. This should help to derive some important socio-economic insights, such as examining the stability of household income and education outcomes of pupils in the household. In the medium to longer term, as the PPMD is extended further and pupils who are currently in school start to progress into post-16 education and beyond into the labour market, it could start to facilitate intergenerational analysis.

The PPMD may help support other departments with policy and delivery priorities through the new analysis. The DfE recognises the potential value of the PPMD across government but may require additional resources to realise these.

## Some PPMD challenges

Due to the scope of the key administrative systems used in the PPMD's creation, some groups within the population are missing. For example, the DfE's school census data, which underpins the National Pupil Database, only collects data about pupils in state schools so it misses some groups of children (for example, independent-school pupils, children who are home-schooled). Other coverage issues relate to Child Benefit data, which is partly used to link parents to their children. Higher-income parents are no longer able to claim Child Benefits, while larger families can only claim for the first 2 children.

There may be solutions to overcome some of these gaps. For example, one tactical solution to overcome the missing children issue would be to match the PPMD to the 2021 population census. However, this would only be a short-term fix as the census is only carried out once a decade, therefore other solutions would be needed to plug the gap between censuses. Including privately educated pupils may also be challenging – Higher Education Statistics Agency data, or names and addresses associated with public qualifications, may help but would need significant time to create reliable linked data.

Another issue is the geographical coverage of the PPMD, which only covers England. Significant development work would be needed to incorporate data from other countries to make it into a UK-wide household data set. This may not be easy for some of the devolved policy areas such as education, where systems vary significantly across countries in the UK. Harmonising data to allow for a consistent comparison across the 4 countries would be valuable, but it would be a considerable challenge to achieve this.

Several participants who were interviewed in this study noted that the significant time and resource investment needed can be a key barrier to development projects like the PPMD being successful. At the time of participant interviews, it was apparent that the DfE analytical team leading on the PPMD development work faced several other competing objectives. Having greater dedicated resources and ministerial backing will improve the chances of the PPMD development work programme being completed and the benefits being realised.

Annex 1 provides a case study which explores further how a household data set could be created to assess intergenerational mobility. It also examines the PPMD in more detail.

By investing in the creation of an administrative household-level data set, analysis which was previously not possible due to limited sample sizes or missing data may become feasible. This has the potential to provide a significantly more informed picture of the current state of socio-economic disadvantage in the UK and crucially how it varies by place and interactions with both protected characteristics and socio-economic background.

### **Recommendation 3**

A UK household data set based on administrative data would be very desirable. The government should set up a central programme to work towards developing such a data set. This should make use of existing data developments like the PPMD, if possible, to get something which can be used across government and beyond immediately. The programme should have an agreed, prioritised plan for further development to serve a wider set of user needs and move us towards our ambition to have a full UK household data set based on administrative data. Critically, a household data set that included children would allow much more accurate targeting of education policy than the current deprivation measure, based on free school meal eligibility.

## 5. Developing better measures of socio-economic background

---

A challenge with measuring disadvantage is that the concept itself is not consistently defined. A child's free school meals (FSM) eligibility is currently a commonly used proxy for disadvantage. However, there are several limitations with this measure which are discussed in more detail in the next section. A key challenge with using a metric for disadvantage is to avoid grouping individuals into broad categories, particularly measures which are binary (resulting in only 2 categories). This is because conducting comparisons between broad categories risks missing out on a fuller distributional picture and instead encourages a simplistic binary view, with no sense of the variation within each group. The geographical analysis of mobility, for example, is severely compromised by the FSM indicator, since the heterogeneity in the non-FSM group is likely to vary greatly across regions. To enable better conclusions on the socio-economic disadvantage, we must move away from binary indicators to more nuanced measures of socio-economic background.

The development of a new longitudinal UK household data set based on administrative data, which contains information about the income and occupation of household members should facilitate the undertaking of new research to explore other potential measures of disadvantage, which are more effective than existing proxies. For example rather than using the FSM eligibility as a proxy for disadvantage, it may be possible to identify the number of pupils in each school living in households whose income is below a set threshold. It may be possible to go further and identify low-income households which take into account the size and composition of the household.<sup>21</sup> It may be feasible to add further dimensions to households in need. For example, identifying households who face other challenges, such as where the main earner is in irregular work or has a lumpy or occasional income, who might also face hardship and need support.<sup>22</sup> By linking to administrative data from other parts of government, it might be possible to research new

---

<sup>21</sup> This method is known as equivalisation. See Wikipedia, '[Equivalisation](https://en.wikipedia.org/wiki/Equivalisation)'. Published on EN.WIKIPEDIA.ORG.



measures which take into account a wider range of factors, such as families who are in temporary accommodation or where health or social care issues are affecting either a pupil or a household member.

## **Weaknesses of current measures of disadvantage**

Being eligible for FSM or having been eligible in the past 6 years (FSM6) are long-established measures for socio-economic disadvantage. However, they do not capture all of those pupils who live in low-income households who may benefit from support or may be considered as experiencing socio-economic inequalities. For example, nearly 2 million children live in relative poverty (after housing costs are considered) but are not eligible for FSM, which has a much lower income eligibility threshold.<sup>23</sup> This could be because of a number of reasons, ranging from restrictive eligibility criteria to stigma issues for the family.

Another well-established set of measures is the Department for Levelling Up, Housing and Communities' (DLUHC) Indices of Multiple Deprivation (IMD), which are place-based measures. These use several data sources, primarily administrative data, to classify the relative deprivation of small geographic areas. The 7 domains of deprivation are then weighted differently and combined into a single score of deprivation. Each of the scores for the small geographic areas can then be ranked to identify their relative ranking. The main issue with the IMD is that while it may provide information about the level of deprivation in one area relative to other areas, it does not provide information about the characteristics of households in that area. Therefore, this may result in missing poorer households who live in more affluent areas and vice versa.

Departments want to understand how to best define socio-economic deprivation and the extent to which FSM is the best proxy for this to allow them to better measure change in

---

<sup>22</sup> The Index of Multiple Deprivation Income domain does something similar to this. See GOV.UK, '[English indices of deprivation 2019: technical report](#)', 2019. Published on GOV.UK.

<sup>23</sup> In January 2021, 1.74 million pupils were eligible for free school meals (FSM). See GOV.UK, '[Schools, pupils and their characteristics](#)', 2022. Published on GOV.UK. However, according to the Department for Work and Pensions (DWP), national statistics on relative child poverty after housing costs are considered there are 3.6 million children living in relative poverty. See DWP, '[Households below average income: for financial years ending 1995 to 2020](#)', 2021. Published on GOV.UK. This is a difference of 1.86 million children who are not classed as eligible for FSM but are classed as being in relative poverty.

socio-economic status and target effective support. It was noted in discussions with the research participants that FSM and FSM6 are binary measures – that is, a family is either eligible or not, largely depending on what side of the very low Universal Credit (UC) income threshold they fall. However, there are many families which may have incomes beyond the UC income threshold who still need support but cannot be identified as they are grouped with families who have much higher incomes. Sometimes, other measures were also used alongside or instead of FSM, such as the IMD, the employment status of parents and whether an individual has been in care.

Some work has already been undertaken to understand the role of FSM and FSM6 as a proxy for socio-economic disadvantage. The Education Data Lab has shown there are differences between continuous FSM and those who drop in or out of eligibility.<sup>24</sup> Research by the University College of London's Social Research Institute found that eligibility for FSM (averaged over the time a child has spent at school) is the best available proxy for childhood poverty, but is of limited use to researchers wanting to understand how key outcomes differ between young people from low, average and high socio-economic backgrounds. It also found that by combining individual and area-level socio-economic proxies into a single continuous index, administrative data can be used to produce robust estimates of family income differences in key educational outcomes.<sup>25</sup>

When discussing place-based proxies, the DLUHC said that there are limited opportunities to amend the IMD as it is a designated National Statistic which is published every 3 to 5 years. From previous consultations, users say they most value consistency in method and definition across iterations. However, a fresh consultation is likely to form part of any future update. The DLUHC has also been considering a new index of opportunity to assess the relationship between place and outcomes. This means they are comparing different people with similar metrics such as FSM, parental education, and parental income.

---

<sup>24</sup> Education Data Lab, '[Long-term disadvantage, part one: Challenges and successes](#)', 2017. Published on [FTEDUCATIONDATALAB.ORG.UK](http://FTEDUCATIONDATALAB.ORG.UK).

<sup>25</sup> UCL, Social Research Institute, '[Measuring socio-economic background using administrative data. What is the best proxy available?](#)', 2020. Published on [ECONPAPERS.REPEC.ORG](http://ECONPAPERS.REPEC.ORG).

## **Issues with identifying persistently disadvantaged pupils**

Due to UC transitional arrangements, it is no longer possible to identify the length of time a pupil has been disadvantaged. Ordinarily, a pupil who is eligible for FSM would not remain so if their family income improved and they no longer met the eligibility criteria. However, to smooth the rolling out of UC, the government introduced some transitional arrangements. Under these, any pupils who are eligible for FSM at any point between April 2018 until the end of the UC roll-out (which is set to be summer 2023 at the earliest) will retain their FSM eligibility for this whole period and until their phase (primary or secondary) of education ends. This applies even if their family circumstances improve and they would ordinarily no longer have been eligible for FSM.

In general, the longer a pupil has been disadvantaged, the lower their average attainment, so it is important to be able to identify how long a pupil has been disadvantaged. This data may already be held by or can be derived from information held in the Department for Work and Pensions (DWP). Rather than ask schools to collect this information from parents, the government should investigate whether changes in legislation in recent years, such as the Digital Economy Act (2017), could be used to identify and share this information with the Department for Education (DfE) and schools, thereby removing the current burdensome process used to collect information on FSM eligibility within schools.

## **Creating and adapting measures of disadvantage in administrative data**

Our discussions identified that creating formal measures for socio-economic disadvantage in administrative data would require establishing an agreed definition or set of attributes for an individual and their household, coupled with measures to indicate what would constitute social mobility. This would cut across departments giving them a common set of indicators to collect (such as parental education levels) rather than relying on proxies which people or households may not fit into.

As new disadvantage measures are developed and tested, they could be added to the proposed strategic UK household-level data set and shared with other government departments. As each department would then be able to identify the same population of disadvantaged households, it may be possible to develop a more rounded view of their circumstances and the level of support they are receiving across the government and beyond. This may help to further increase our understanding of the causes of

socio-economic disadvantage and to devise more joined-up policies and practices to tackle these issues.

Annex 2 provides a case study which explores establishing agreed proxy measures for socio-economic disadvantage and social mobility in further detail.

#### **Recommendation 4**

It is important to collect information to identify the length of time a pupil has been disadvantaged as attainment outcomes tend to be lower on average for the more persistently disadvantaged. The DfE should undertake a review to identify how best to collect this information, bearing in mind it may already be held by or can be derived from information held in the DWP.

#### **Recommendation 5**

The Office for National Statistics, with input from the Social Mobility Commission, should lead a cross-government project to conduct research into developing improved social background measures to help form a wider picture of the full range of socio-economic characteristics experienced across the population. These should be developed for the use of both policy and operational purposes. These could be shared across government, enabling policymakers in all departments to track households from a wider range of socio-economic backgrounds in a joined-up way, and to better target support.

## 6. Filling important data gaps

---

Throughout the series of interviews and the development of the new Social Mobility Index, several key data gaps have been identified. These range from the capture of some characteristics of interest (such as wealth), to the ability to compare variables consistently across the UK (data harmonisation). Also the limitations of small sample sizes prevent analysis of the interactions between protected characteristics, region and socio-economic background. These gaps act to limit the extent to which analysis can inform the current picture of socio-economic disadvantage and make tracking progress in targeted policy interventions and its implications for social mobility even more challenging. The UK Statistics Authority's Inclusive Data Taskforce, led by Dame Moira Gibb, has already recommended that data on socio-economic background should be included in administrative data collection, which we welcome.<sup>26</sup> In this chapter, we discuss some of the data gaps that currently exist.

### Occupational data

One of the more commonly used indicators of socio-economic position is the social class of an individual as measured by their most recent occupation. But occupational data is not routinely collected by any part of the government as part of carrying out its administrative functions. This is a key gap in the data that would ideally be available for improving our understanding of socio-economic disadvantage.

The best option currently available is the Labour Force Survey (LFS), which although of high quality, does not allow for very detailed geographical breakdowns due to sample-size limitations. Only administrative data could allow for the meaningful analysis below local authority level. HM Revenue and Customs (HMRC) is also currently consulting on adding Standard Occupational Codes (SOC) to the data that it collects. If this is implemented, and more critically, if the data is shared, then this could greatly improve the situation.

---

<sup>26</sup> UK Statistics Authority, '[Inclusive Data Taskforce recommendations report: Leaving no one behind – How can we be more inclusive in our data?](#)', 2021. Published on UKSA.STATISTICSAUTHORITY.GOV.UK.

Another possibility is to make use of occupational data which is collected in the decennial UK population census. By linking this to education and earnings data, we could measure the relationship between the occupations of a child's parents relative to its outcomes. Naturally, this would just provide a snapshot based on a person's occupation at the time of the census once every 10 years. This would nonetheless be useful for research purposes and the Department for Education (DfE) should consider adding this to its Pupil Parent Matched Data (PPMD) as part of its ongoing development. The Digital Economy Act (DEA) could provide the necessary legislative cover to enable this to happen.

The Office for National Statistics (ONS) has also developed a data set which takes a one per cent sample of census records and has linked them together longitudinally across successive censuses, potentially allowing for better mobility analysis.

While linking the population census to the PPMD would go some way to filling the parental occupation data gap, in an ideal world, occupational data would be collected routinely by the government as part of one of its administrative functions. This would help researchers to explore questions such as how changes in parental occupation impact an individual's or household's socio-economic status and whether the impact is temporary or permanent. This could potentially help inform policy development by enabling an improved picture of how a policy is targeted and measuring the impact it may have on social mobility.

The collection of administrative occupational data could have wider benefits, especially if used for operational purposes. For example, during the pandemic, it may have been helpful to identify which occupations were worse affected or at risk of losing their jobs, and targeting appropriate help.

Therefore, the government should implement current proposals to collect and – critically – share occupational data as part of an administrative function. One potential method could be if the HMRC collected this information from employers through the pay-as-you-earn (PAYE) systems for workers and the self-assessment tax return for the self-employed.

## **Parental and family income**

In many jurisdictions, the tax records of parents and children are linked. This has enabled detailed and high-quality analysis of income mobility, such that carried out in the US by

Chetty and others.<sup>27</sup> Unfortunately, there is no equivalent linking in the UK. Even worse, the second-best option of using the LFS is not available (as it is for occupation), because the LFS does not ask about parental income. This means that there is no official data source for the rigorous analysis of income mobility. We have to rely instead on academic panel studies, which suffer from major limitations (irregularity in the birth cohorts chosen, along with high attrition and lack of representativeness).

Therefore, the government should consider how such family linking might be feasible in the UK, to ensure that the UK does not fall behind other countries in its ability to analyse income mobility trends.

## **Birth cohort studies**

The UK has 4 main birth cohort studies: the National Survey of Health and Development, established in 1946, the National Child Development Study, established in 1958, the 1970 British Cohort Study, and the Millennium Cohort Study, established in 2000. Despite the limitations of such studies they can provide very valuable insights into social mobility and socio-economic diversity. Many of these insights would not be available even from improved administrative data. For example, birth cohort studies can give very detailed information on parenting, families, and the conditions of childhood.

There are new longitudinal studies, although not full birth-cohort studies, that have recently been commissioned: the Early Life Cohort Feasibility Study, Children of the 2020s, and the COVID Social Mobility and Opportunities Study. These studies “will follow groups of young people whose formative years have been shaped by the COVID-19 pandemic and other recent world events.”<sup>28</sup>

We welcome UKRI’s endorsement of a new UK-wide birth-cohort study, and encourage the government to support this work, and future such studies.

---

<sup>27</sup> Raj Chetty and others, [‘The fading American dream: trends in absolute income mobility since 1940’](#), 2017. Published on PUBMED.GOV.

<sup>28</sup> UCL, [‘Our studies’](#). Published on CLS.UCL.AC.UK.

## **The Labour Force Survey**

The LFS has proved to be a valuable resource for the analysis of social mobility. It is a well-conducted survey with a very large sample size, enabling regional breakdowns of social-mobility measures, as well as breakdowns by protected characteristics. Social Mobility Commission (SMC) analysts who have used both data sources also reported that the quality of the occupational data collected through the LFS has tended to be higher than in the Census, because the LFS is conducted by trained interviewers.

While there are inevitably many items that compete for space on a survey, the government should consider whether further items which are relevant for a better understanding of social mobility could be added to the LFS. For example, parental education, wealth, housing, or social capital. These would allow a much more detailed and reliable analysis of social mobility.

## **Trajectories for income and work, situational changes, and material deprivation**

Participants noted that there is a mixed availability of 'softer' data to help understand an individual's context and economic activity. Examples included data on non-earned income or wealth, hours worked, a change in household circumstances, and a child's non-cognitive skill levels, all of which may provide a broader understanding of household circumstances. Consequently, while some analysis can be undertaken to understand the links between socio-economic status and income and employment outcomes, contributing factors cannot always be explored fully. This limits the government's ability to understand where limited resources are best targeted for maximum impact and value for money, and where there are incentives (or disincentives) to engage with the system.

There are also other variables which may be important in terms of understanding historic or future mobility outcomes. For example, these could include factors such as the number of hours worked by parents as longer working hours may impact their ability to oversee their children's school work or the childcare arrangements for pre-school children which could impact the development of younger children. However, further research would be needed to improve our understanding of which other factors could be of greatest help in



better understanding the drivers of social mobility, and to identify which new data is worth collecting regularly.

Where respondents knew the PPMD, they believed it could be key to understanding the wider context and softer information through its integration with other data sources from the Department for Levelling Up, Housing and Communities (DLUHC), health data from the National Health Service (NHS), information on interactions with the criminal justice system from the Ministry of Justice (MoJ), and survey data such as the Longitudinal Study of Young People in England (LSYPE) and Millennium Cohort Study.

It was also understood that introducing additional variables to be collected in some contexts may provide further insight to enable the investigation of intergenerational social mobility and the contribution of internal migration and geographic mobility. Such variables could include adding data on occupation and some geographic information about a location of home and work (subject to this not undermining data security) to key data sets such as the Longitudinal Education Outcomes (LEO).

As well as having information about the level of resources that individuals and households have, developing a measure or set of measures about the sufficiency of resources is also really important to help identify when people are getting into trouble. For example, information about a person's debt and their lack of ability to both eat and heat their households adequately are all very important. This is especially true during times when the cost of living is rising sharply and levels of support are not keeping up, forcing some people and households to make tough choices, and rely on foodbanks.

## **Ethnicity and other equality, diversity and inclusion data**

It was observed that the detail and accuracy of ethnicity data collected by different departments and in different data sets are mixed. The Race Disparities Audit (2017) highlighted that administrative data can suffer high levels of non-recording of ethnicity and overuse of 'other' categories, undermining the ability to identify differences in how people in each ethnic group are treated.<sup>29</sup> Absence or incomplete ethnicity data may skew our understanding of the take-up of support among different groups who may not wish to

---

<sup>29</sup> GOV.UK. '[Race disparity audit](#)', 2017. Published on GOV.UK.

provide this data, as well as what works for these groups, and the impact that policies and changes have for different ethnicity groups.

The decennial UK population census data could be used as a tactical solution to fill gaps in some administrative data sources that are used for analytical purposes if permissions are agreed. If successfully linked, it may be possible to use census data to validate existing ethnicity data in data sets where data quality may be of concern. Finally, it might be possible to add more detailed information about ethnicity to some data sets. For example, where a data set only collects this information at the 5 broad ethnic categories (Asian, Black, Mixed, White, Other), it might be possible to use the population census to provide a more detailed breakdown at the '18 + 1' ethnic group level.

## **Harmonised education data**

As education is a devolved policy area, each nation can shape its own qualification system. Each devolved nation's government has influence over which data it decides to collect and which education statistics are published. As a result, it can be difficult to consistently compare educational attainment measures such as attainment at age 16 across the UK, such as between England and Scotland.

A lack of harmonised education data across the UK for the key indicators of the index limits the ability to make comparisons and benchmarks between UK nations and regions.

The National Statistician, working in partnership with the devolved administrations, could identify the key data gaps which are hindering our understanding of socio-economic inequities and the impact on key policy areas, and how they could be addressed.

There is also an important problem with measures of higher education, which are centrally collected by the Universities and Colleges Admissions Service and Higher Education Statistics Agency. These bodies do not report relationships with standard measures of socio-economic background but use measures such as the Participation of Local Areas, which are area-based and involve major problems of interpretation and comparison with standard socio-economic background measures.

## Parental education

Similarly to income, the analysis of education mobility is severely hampered by the lack of good-quality administrative data on parental education. Linked data on parents' and children's education would greatly improve this.

Pupil-level data started to be collected by the DfE in 2002. If the PPMD was extended backwards and was longitudinal from 2002, some of the parents with children in the school system today would have been pupils themselves in that time. Therefore, we would have their education data, and so with each passing year, we would have parental education data for an increasing proportion of the school population. However, to reach its full potential, the data set needs to be further developed to include SOC codes, hours worked, and ultimately pupils who were educated in private schools, to give a more accurate picture.

### **Recommendation 6**

The government should implement current proposals to collect and – critically – share occupational data which is needed for a better understanding of social mobility and the causes of socio-economic disadvantage. This could be collected by HMRC from employers through the PAYE and national insurance system. If occupational data is not collected, any household data set would be lacking critical information.

### **Recommendation 7**

Linked administrative income and education data between parents and children is a key evidence gap for social-mobility analysis. Without this, researchers cannot look at the earnings or education of today's adults and compare them with the earnings or education of their parents. There is currently no clear way forward to achieving this in the UK. The National Statistician should work with the SMC and other departments to identify and assess potential solutions for developing administrative data systems, and to put in place a programme of work to address this important evidence gap.

### **Recommendation 8**

We welcome UKRI's endorsement of a new UK-wide birth cohort study as a successor to the 4 studies that have taken place since 1945. The government should consider how this critical work can be best supported now and in the future, so that further large gaps do not arise.

### **Recommendation 9**

The government should commission the National Statistician to lead a cross-government review of UK nations and in partnership with the devolved administrations. This would identify the key data gaps that are hindering our understanding of social mobility and how they could be addressed.<sup>30</sup> This review should take both administrative data and survey data (such as the LFS and the WAS) into consideration.

---

<sup>30</sup> The National Statistician the Chief Executive of the UK Statistics Authority, Permanent Secretary of the Office for National Statistics and Head of the Government Statistical Service.

## 7. Process and practicality issues

---

Several other process and practical issues emerged during this study. If addressed these could contribute to an improved understanding of the causes of low social mobility and socio-economic disadvantage. These are summarised below.

### **Dedicated analyst resource**

It was noted during this study that it was sometimes difficult to dedicate scarce analyst time to address some of the wide-ranging and bigger policy questions that exist. To increase the potential resources available to help address longstanding research questions, any new linked administrative data sources and household-level data sets must be accessible to government analysts, academics and third-party researchers safely and securely. The latter may have the capacity to provide more dedicated and focused time, which is needed to address difficult and detailed research questions. Academics and researchers may also bring different research insights which could complement the efforts of government analysts.

### **Improved knowledge of what data exists**

It was reported that there was a lack of knowledge about what data exists across the government estate which might be useful in providing some new insights on socio-economic inequalities. There was also a vagueness among officials in one department about what might be happening in other government departments in terms of developments, which could be useful to them. Therefore, improved communication between departments regarding the data available and developments on data sets of interest may yield wider benefits across the government. A centrally-held and – critically – regularly updated list of available data would go a long way to improving this.

### **Empowering the ‘head of data’ role**

Another issue was that sometimes heads of data in departments had difficulties making the case for changing or collecting a new variable through their administrative processes. Analysts generally have a good idea about the data needed to answer key questions and

provide important new policy insights. However, they may lack the influence needed to bring data changes about, as these may require ministerial and wider departmental support to be approved, especially where collecting the new data would require changes to be made to administrative and IT systems. It would help if the heads of data and statistics in government departments had a greater influence in the prioritisation and discussions regarding changes to administrative systems.

## **Strengthening the oversight of social mobility**

Primarily we are focusing on identifying data gaps which affect our understanding of the factors that help to generate better social mobility outcomes. But policy responsibility for addressing aspects of social mobility, be it directly or indirectly, is dispersed across government and not always joined up. For example, the Department for Education is responsible for addressing the gap in education outcomes between disadvantaged pupils and their peers, the Department for Work and Pensions focuses on tackling child poverty and getting people into work, the Department for Health and Social Care is responsible for addressing differences in health outcomes which lower-income families face, the Department for Levelling Up, Housing and Communities (DLUHC) has overall responsibility for troubled families and place-based outcomes, and HM Treasury has responsibility for income inequity across the income distribution. A number of these areas are dependencies for other policy areas or are interlinked, but there does not always seem to be much joining-up. Greater progress might be achieved if this important plan was led by a specific minister and department with overall responsibility across the government for improving social mobility and reducing socio-economic differences. This could provide greater clarity about how each department is contributing to this overarching objective in terms of delivering their goals and facilitating other parts of government to achieve theirs.

## **Recommendation 10**

As an arms-length body, the Social Mobility Commission does not have the power to require government act on the recommendations it makes. Therefore, we encourage the government to assign the recommendations in this report to an appropriate central body that can take coordinated action on them. For example, the Central Digital and Data Office responsible for the National Data Strategy or the ONS Centre for Equalities and Inclusion.

In addition to the previous recommendations, the appointed lead should also be responsible for publishing a regular audit of the non-sensitive data sets held in each department (although the release of the actual data, where appropriate should be done in a properly planned way, and not necessarily as part of this audit).

To help deliver on these recommendations, the government should also give a relevant department such as the Cabinet Office or the Department for Levelling Up, Housing and Communities' overall responsibility for oversight of work across government to address socio-economic disadvantage. This would encourage departments to set meaningful objectives on social mobility data gaps and social mobility more generally, and hold departments to account for delivery.

# Conclusion

---

To conclude, socio-economic data has the potential to significantly enhance policy-making targeted at improving social mobility in the UK. However, currently, there exist major data gaps and barriers to building the appropriate data sets required to achieve this. This report sets out an aspiration for what the socio-economic data environment could be like in the future. It also sheds light on some of these barriers and data gaps, and hopes to stimulate debate and invite the government to take action to improve the socio-economic data available.



# Annex 1: Case study – a household data set to measure social mobility

---

## Overview of case study

In this case study, we examine one of the main gaps identified by policy officials and analysts who were interviewed as part of this research, namely the need to create an administrative data-based UK household-level data set. Linking together administrative data sources across government, including education, employment, health and justice data would enable much more detailed analyses of intergenerational social mobility and socio-economic inequalities at lower levels of geography. As well as exploring the rationale for doing this, we focus on whether the Pupil Parent Matched Data (PPMD), which is an initiative currently underway in the Department for Education (DfE), could be developed further and expanded beyond England to become this flagship UK household data set.

## Description of the problem

Current administrative and survey data do not offer a consistent and high-quality way to link household members together to understand the different factors that may contribute to income, behaviour, and outcomes.

Researchers across different departments want to understand more about socio-economic disadvantage through the wider characteristics of the household. For example, whether it is a dual or single-parent household, what the household income is, levels of parental education, and parental occupation.

Many current administrative data sources collect information at an individual level (for example, HM Revenue and Customs' (HMRC) earnings data, and DfE's pupil data). Others are reliant on questions being 'built-in' to understand family context, meaning they may lack information or have incorrect or out-of-date information. In some instances, there are also limitations on what departments can collect, both in terms of their remit and purpose of data collection, and in the time available to capture this for example, through limited survey lengths, or the customer experience while providing this information.

In addition, some administrative data sources are often collected at one fixed point (for example, Child Benefit data) and do not capture how households may move and change over time, which can impact individual outcomes.

Departmental representatives wanted to explore how different data sets could be brought together to provide an overview of different households, and then explore how these data can be linked to help provide a better understanding of socio-economic disadvantage and intergenerational social mobility.

## **Why is this needed?**

A lack of understanding of household composition and changes to this, as well as a lack of information about the needs and experiences of other household members, can impact our understanding of socio-economic disadvantage and intergenerational social mobility. For example, when looking at key stage 4 and key stage 5 outcomes, the DfE can look at the prior attainment of pupils, their ethnicity and where they live using administrative data, but it does not have much data on parental occupation or household income. These are important independent variables which will have an impact on socio-economic disadvantage.

Having further information on the household, alongside linking this to other data such as educational outcomes, parental employment, and health data, will help departments to better understand the drivers of social mobility. In a practical sense, this data should help to provide better insights into the impact that different policies have on wider household members and improve the effectiveness of support programmes.

## **What will it improve?**

Improved linking between sources to improve household-level data could help in understanding a range of issues for departments, in particular the interaction between health and educational outcomes; the impact that interaction with the criminal justice system can have on individuals and their households; and an understanding of how different early years or place-based interventions can impact educational and occupational outcomes.

In time, with better-linked household data, further exploration can be undertaken on these issues to better understand intergenerational social mobility.

## **How can this be achieved?**

Much of the research to date on understanding intergenerational social mobility and socio-economic disadvantage has been based on longitudinal surveys and cross-sectional surveys such as the British Cohort Study or the households below average incomes data set, which is based on the Department for Work and Pensions' (DWP) Family Resources Survey. While researchers can collect richer data in surveys, sample sizes tend to limit the amount of analysis that is possible, for example, geographic area-based analyses of socio-economic disadvantage.

For these reasons, there has been an increasing desire by departments to link together administrative data sources so they can better understand and address socio-economic differences in different areas or to evaluate policy. However, several of the linked data sets which have been successfully created to date can only be used to answer specific research questions, due to issues with establishing data-sharing arrangements across departments which are both complex and lengthy to arrange, and concerns about data security. For example, the data set created by the Supporting Families team in the Department for Levelling Up, Housing and Communities is focused on understanding the impact of the programme and can only be accessed by those working on evaluation. This reduces the potential utility of the data.

Notwithstanding the data sharing challenges and data security concerns, there are some data set developments where progress is made in terms of creating a household-level data set which can be used for exploring a wider set of research purposes. One of the most promising initiatives is how the DfE is developing its PPMD.

This work initially started as a result of the 'schools that work for everyone' consultation response, where the government said it wanted to identify pupils who are not captured by the disadvantage measures used in education, but whose families were just managing to get by.<sup>31</sup> The DfE set up a feasibility study to link personal characteristics of pupils and

---

<sup>31</sup> Department for Education, 'Schools that work for everyone', 2016. Published on GOV.UK.

school level and attainment information to parents' and guardians' tax credits information, pay, and benefits data from HMRC and the DWP.

As well as creating the household structure for pupils in state education, this has enabled the DfE to link parental and guardian income data to pupil education outcomes for the first time.

The DfE has plans to develop the PPMD further, including: working with HMRC and the DWP to develop a more reliable metric for household income which plugs some of the gaps in the data set; expanding the single-year proof of concept data set into a longitudinal data set, which would allow analysts to better understand how circumstances and outcomes have changed over time; linking the PPMD to the Longitudinal Education Outcomes (LEO) data set, which would substantially enrich the data source to cover post-16 outcomes; and linking the PPMD to survey data, for example, the Longitudinal Survey of Young People in Education (LSYPE) and the Millennium Cohort Study.<sup>32</sup>

Beyond this, there is potential to further develop the PPMD so it can be used to support a wider set of users across government as well as externally to academia and research organisations. For example, by linking health data to education outcomes, we can answer questions such as how do health issues during childhood impact education outcomes.

The DfE should proactively promote the work it is doing across government so that other departments are aware of this opportunity and can explore whether their needs could also be catered for, to achieve a wider set of benefits. By doing so, this may help the DfE to achieve buy-in for accessing other data sources to link into the PPMD. We could potentially offer to help DfE promote PPMD across departments and provide additional resources to support the development work so that the benefits that this data source can provide is reaped.

Work should also be undertaken to critically appraise whether the PPMD could be developed strategically to become the new strategic UK household-level data set.

---

<sup>32</sup> Currently Longitudinal Education Outcomes (LEO) only include adults up to around age 35, so there are currently relatively few with children that are reaching the end of their school education. However, as LEO is extended, there will be an increasing number of parents who have children that have completed their education and moved into the labour market, which will start to enable researchers to use it for analysing intergeneration mobility.

Participants in this research study identified this as the top data gap that needs addressing.

## Enablers and challenges

Many enablers and challenges were highlighted which would need to be considered to ensure the effectiveness and quality of options for data linking, and establish protocols for their ongoing use both by government departments and wider partners. Challenges fell into 3 key areas, including:



### **Data sharing and data security concerns**

Concerns within and outside of government over the implications of sharing and linking data, as well as the risks of the data inadvertently being disclosed by a third-party tends to generate caution and inhibit progress on sharing data across government and with third parties. Also data-sharing agreements are complex and generally take a disproportionate amount of time and resources to set up. The DfE has used the Small Business, Enterprise and Employment Act (2015) legislation which enabled the creation of the LEO data set to do the data linking needed for PPMD. However, it is likely to need other data-sharing arrangements to be put in place to bring in data from other parts of the government. The relatively new Digital Economy Act 2017 could help to facilitate this. The relevant departmental ministers must make this a priority, as this is much more likely to ensure this happens.



### **Concerns relating to the time and resources needed to develop and embed appropriate solutions such as the PPMD, and to manage and respond to requests once they have been established**

There was a view that data-linking approaches needed to stem from a long-term vision and commitment from different departments demonstrated through effective communication of their aims, purpose, and potential; and through suitable levels of resourcing for delivery. It was recognised that for data linking to have an impact, clear budgets and resources need to be identified and protected for an extended period so the appropriate systems can become embedded, promoted, and used effectively.



## **Barriers to data linking offering the analytical opportunities needed**

If the PPMD is seen as having a key role in addressing data linking gaps, it is crucial that it collects the fields of interest for different departments, and that opportunities to match with different administrative and survey data are explored, while remaining within the DfE's vision for the data set. In addition, barriers and delays relating to data sharing must not delay and limit the overall roll-out and use of PPMD.

To address these challenges, partnership working is needed with DfE to understand where offers of support may be welcomed regarding the business case and development of the PPMD. This includes political support to make sure its value is recognised and sufficiently realised, and provide resources to support this. This engagement could also be used to help explore more detailed feasibility of how measures to better understand socio-economic disadvantage and causal factors in social mobility can be included and explored.

In addition, sufficient resources and expertise need to be established to address concerns around data sharing. This will promote best practice around the development of data-sharing agreements – learning from lessons during the pandemic where data sharing was seen as being expedited.

# Annex 2: Case study – proxies for disadvantage and social mobility

---

## Overview of case study

In this case study, we explore how government departments might better understand the factors influencing socio-economic deprivation and social mobility, and the role different measures play in acting as an effective estimate for this. Through the use of effective data-sharing measures and data linking, departments will be better able to determine the key indicators of socio-economic deprivation and social mobility. This will allow them to better measure change in socio-economic status and effectively develop and target support.

## Description of the problem

There are not any well-established, overarching, robust measures about an individual's or family's socio-economic position on most administrative data sources. As such, policy officials, practitioners and analysts in government departments, and academics and researchers that use administrative data tend to rely upon a range of proxy indicators of socio-economic status.

These include:

- free school meal (FSM) eligible and FSM6 claimants<sup>33</sup>
- the Index of Multiple Deprivation (IMD)
- the Income Deprivation Affecting Children Index which measures the proportion of all children aged 0 to 15 years living in income-deprived families
- the Participation of Local Areas (POLAR) measure, classifies local areas across the UK according to the young participation rate in higher education
- children living in absolute or relative poverty.

---

<sup>33</sup> Being eligible for FSM or having been eligible in the past 6 years (FSM6) are long-established proxy measures for socio-economic disadvantage.

As well as there being a range of measures in use, different ones are sometimes used across a department's areas of responsibility. For example, within education, FSM and FSM6 tend to be used in schools (for example, calculating pupil premium entitlement), IMD is used in apprenticeships, and POLAR is used in higher education. This makes it difficult to track what is happening across different stages of education.

It is also acknowledged that these proxy measures may not capture all of those who may be considered as experiencing or having experienced socio-economic disadvantage.<sup>34</sup> There are a variety of reasons. For example, some people who are eligible for FSM do not make a claim, while others may not be eligible for FSM but could be struggling and need support.

The absence of one standard, established measure for socio-economic disadvantage means different understandings exist within and between different departments as to what exactly they should be measuring and capturing. Differences in measures also mean that the exploration of contributing factors for socio-economic disadvantage, targeting of support, and allocation of resources may be missing key groups who might benefit from this.

Without a clear definition and measure for socio-economic disadvantage, it is challenging to measure change in an individual's or a household's situation, and understand the contributing factors to social mobility. This makes it harder to monitor people across the system to ensure their life chances are improved.

## **Why is this needed?**

It would be helpful to establish a new measure or set of measures which better identify socio-economic disadvantaged households that could be used more widely across government. This would enable different users to have a consistent way of identifying households and individuals who need support, and give policy and analytical officials a clearer picture of what support is needed, and what works best for different subgroups.

---

<sup>34</sup> UCL, Social Research Institute, '[Measuring socio-economic background using administrative data. What is the best proxy available?](#)', 2020. Published on ECONPAPERS.REPEC.ORG.



From an operational perspective having a clearer understanding of who may be in this group will ensure support can be more effectively targeted and delivered.

## **What will it improve?**

Having a more effective way to identify people and households who have experienced socio-economic disadvantage will more efficiently use the resources that are in place to improve people's life chances. This is from an individual, household and place-based perspective.

Importantly, having these common measures will allow further research and analysis to be undertaken to better understand the factors contributing to social mobility, allowing the more effective development and delivery of support mechanisms.

## **How can this be achieved?**

The main reason why different proxies are used across departments is that the administrative data sources they collect and maintain lack the necessary information for them to create better measures. For example, the DfE's school census collects details about each pupil in the state education sector each term, but this does not collect data about a pupil's parents or siblings, or household income. As such, the DfE does not have a household-level data set with which to identify which households have low incomes and has to rely on the number of pupils who are eligible for FSM.

However, as noted in case study 1, there is a desire among policy officials and analysts to create a household-level data set which links together a range of key data from across the government, including parental employment, income, education and health details. The development of a new strategic UK household-level data set could facilitate the development of new alternative measures of disadvantage. This in turn will provide a more accurate picture of households who are facing socio-economic disadvantage. These will provide more insights about which households need help and how best to provide support.

Once a new measure or set of measures is developed, data-sharing agreements could be set up with departments so that these new measures could be linked into their key administrative data sets where possible. This would help departments to focus on the same group of disadvantaged households, which could facilitate a consistent, joined-up approach across government.

Some departments may also want to use the new measure or set of measures for operational purposes. For example, to determine pupil premium funding for schools. This might necessitate further data-sharing agreements to be put in place, but there would be a strong rationale to improve the targeting and efficiency of funding provided.

## **Enablers and barriers**

Current approaches to proxy use for socio-economic disadvantage have been developed and used by different departments in response to their core policy objectives and needs, and link directly to the development and delivery of policies and programmes. For some, changes to proxy measures may create concerns regarding their IT systems and the impact on comparability with previous data.

The development of any measures would need to explore where flexibility can be built in to support their use in different circumstances or when working with different specialist groups to share examples of what proxies work best. In addition, these examples should include options to ensure existing data can still be used to compare with previous cohort data.

To achieve such potentially complex changes, it will be crucial to engage across government more widely to allow input and support into data options, identify challenges for operationalising these, and ensure buy-in based on the benefits this might achieve through more efficient and effective use of resources. There is interest from different departments to do better in terms of capturing and utilising data to better understand and support social mobility, but further support on 'best practice' ways to do this is welcomed.

# Glossary

---

## Executive summary

LFS	Labour Force Survey
SMC	Social Mobility Commission
FSM	Free school meals
NFER	National Foundation for Educational Research
ONS	Office for National Statistics
IDS	Integrated Data Service
DEA	Digital Economy Act 2017
UC	Universal Credit
PPMD	Pupil Parent Matched Data
DfE	Department for Education
WAS	Wealth and Assets Survey
HMRC	His Majesty's Revenue and Customs
DWP	Department for Work and Pensions
CDDO	Central Digital and Data Office

## Chapter 1

OECD	Organisation for Economic Co-operation and Development
NCDS	National Child Development Study
BCS	British Cohort Study – a birth-cohort study that follows the lives of more than 17,000 babies born in England, Scotland and Wales in a single week in 1970.
MCS	Millennium Cohort Study – a birth-cohort study that follows the lives of about 18,818 babies born in the UK in 2000–2001.

## Chapter 2

ICO	Information Commissioner's Office
DHSC	Department of Health and Social Care
	Small Business, Enterprise and Employment Act
HESA	Higher Education Statistics Agency
LEO	Longitudinal Education Outcomes
SRS	Secure Research Service

## Chapter 4

DLUHC	Department for Levelling Up, Housing and Communities
MoJ	Ministry of Justice

## Chapter 5

IMD	Indices of Multiple Deprivation
-----	---------------------------------

## Chapter 6

SOC	Standard Occupational Codes
PAYE	Pay as you earn
	Early Life Cohort Feasibility Study – a study of the feasibility of carrying out another birth-cohort study in the UK in the near future.
	Children of the 2020s – a birth-cohort study of babies born in England at the start of the 2020s.
COSMO	COVID Social Mobility and Opportunities Study – a cohort study of more than 12,000 young people in England, who were in Year 11 in the academic year 2020/21, examining the impacts of the COVID-19 pandemic on education and social mobility.
LSYPE	Longitudinal Study of Young People in England
	Race Disparities Audit (2017) – a government review of how people of different ethnicities are treated across public services.