Data and public policy: trying to make social progress blindfolded

A report for the Social Mobility and Child Poverty Commission

CONTENTS

Foreword

Low social mobility and high child poverty will not be tackled unless there is a good understanding of their causes and of the key points at which policy can make a difference alongside an understanding of the effectiveness of different policy interventions.

In other words, the intelligent use of data to develop evidence-based policies and prioritise efforts must be at the heart of strategies to improve social mobility and reduce child poverty.

Academic research, mainly using the British Birth Cohort Studies, was instrumental in highlighting that Britain had a social mobility problem. Political action has followed. Administrative data on the educational progress of children has been instrumental in informing us about the nature of the problems we face but also in providing valuable success stories, such as the huge improvement in educational attainment by poor children in London.

This agenda is even more crucial at a time when resources are scarce. Every major political party is promising to make large spending cuts if they win the General Election in May. Public spending is set to reach historical lows during the next Parliament. Without urgent action to unlock data the risk is that lack of information will frustrate good intentions on social mobility and child poverty, leading policymakers to choose the wrong priorities and use shrinking resources less effectively than they otherwise could. The ultimate cost will be failure to improve the prospects for children from disadvantaged backgrounds.

While some important progress has been made over the last few years in allowing administrative data to be used more effectively – for example, through the work of the ESRC Administrative Data Taskforce and the Social Mobility Transparency Board – it needs to be accelerated.

This report is an important contribution to the debate: it demonstrates how more intelligent use of data can make a difference; it identifies the legal, institutional and cultural barriers that exist to making progress; and it makes a number of specific recommendations for how barriers to using data more intelligently can be removed.

The long term ambition here is to build a system that can track progress of children and young people across a range of outcomes from cradle to adulthood. This would cover school progress, university application and places attained, vocational qualifications, employment and unemployment and earnings. Such an ambition will be a long haul but this report highlights key first steps on this journey.

Professor Paul Gregg
Social Mobility and Child Poverty Commission

Introduction

Successive governments have put improving social mobility and reducing child poverty at the heart of their public policy agendas and have invested much effort and resources in trying to achieve these goals.

But policymakers have been hampered by a lack of information about which policy levers make a long-term difference to children's lives.

Data that can shed light on this is collected by the health service, early years providers, schools, further education colleges, universities and the tax and benefit system but - because of concerns over the implications of sharing and linking data - it is often not used to its full potential.

As a result, policymakers are left trying to make social progress "blindfolded" by limited knowledge about the problems they are trying to address and the potential impact of their decisions.

This report looks at the ways in which data can be better used to inform and guide public policy. In particular, the report focuses on removing **the barriers to sharing 'de-identified' administrative data** [i] Academics and policymakers have identified this as a major block to learning more about social mobility and child poverty.

If more data is shared, we can improve the evidence base on which social mobility and child poverty interventions are based. Greater sharing of data will improve the effectiveness and efficiency of policy interventions and, as a result, reduce public expenditure.[ii] As the next government will face considerable challenges in delivering more effective public services in the face of continued austerity, using the information it collects more intelligently should be high on its agenda.

This report identifies the problems that data sharing currently faces; and it makes 19 recommendations to improve data sharing.

This is by no means the first report to look at data issues and progress has been made in the last few years. In May 2012 the current Government created the Social Mobility Transparency Board; its main aim is to use data more intelligently to develop a richer understanding of social mobility. The Government's Open Data agenda has also helped, in particular through the *Open Data* White Paper and the creation of the Open Data Institute.[iii] In addition, the Cabinet Office is currently working on a Data Sharing Bill.

Beyond Government, the Public Accounts Committee has looked into data issues;[iv] the Law Commission has produced a report on data sharing,[v] and in September 2014 the Royal Statistical Society launched a 'Data Manifesto' in which it described data as the driver of prosperity.[vi]

Dean Machin[vii]
March 2015

Acknowledgements

Data sharing

Administrative data – data that is collected routinely by government and other public bodies in order to fulfil their functions – is extremely valuable to social mobility and child poverty researchers and policy-makers. It can be used to assess the effectiveness of policy interventions and to learn more about the issues being tackled.

Table 1 lists some of the key administrative databases relating to social mobility and child poverty; it illustrates administrative data's public policy value. For instance, an assessment of the long-term effectiveness of early childhood interventions would benefit from access to administrative data at each of the subsequent life-stages. Linking data on early childhood interventions to the relevant HMRC records would help us to better understand the long-term value-for-money of early years investment.

Table 1: Key administrative databases relating to social mobility and child poverty

	Life-stage	Relevant database or body
1	Early years (0-4)	Sure Start On database
2	School	National Pupil Database (NPD)
3	Further education	Individual Learner Records (ILR)
4	The transition from school to university	Universities and Colleges Admissions Service (UCAS)
5	University	Higher Education Statistics Agency (HESA)
6	The transition from university to work	HESA and Destinations of Leavers from Higher Education Survey (DLHE)
7	As individuals' work-lives progress	Her Majesty's Revenue and Customs (HMRC) and The Department of Work and Pensions (DWP)

While a huge amount of administrative data is collected, most of it is not *shared* between different public bodies or with researchers; and it is the exception for different data-sets to be *linked* to each other.

The public benefits

The intelligent use of administrative data will increase our chances of lifting children out of poverty and improving their life-chances. The potential financial savings are also significant. In 1990, Finland started to conduct its census using only administrative data. This has resulted in substantial savings with little demonstrable drop in the quality of the data collected.

> 'The United Nations Statistics Division estimates that the 2001 Finnish 'census' cost just €0.2 per head, compared to an estimated €6.2 per head in the UK'[viii]

In many cases, administrative data is the best data available to researchers and policy makers. The UK Statistics Authority has noted that the use of administrative data 'saves money compared with conventional surveys, and the statistics are often more complete and available for smaller geographical areas… [the statistics] can

also be joined up with records from other sources to give a richer basis for analysis.'[ix]

The potential impact on policy is well-illustrated by the work of Raj Chetty and colleagues who analysed local variations in social mobility in the USA using administrative data from over 40 million individuals, linked to the tax records of their parents. They found substantial variation in intergenerational mobility between different local areas and identified five factors associated with high social mobility. These include lower income inequality, higher social capital and stronger family structures.[x] Similar studies in the UK might transform our understanding of social mobility and so help us design more effective policy interventions.[xi]

Why is data sharing important to social mobility and child poverty?

There are many social mobility and child poverty questions which greater access to administrative data would help answer (Table 2 lists some examples).

To illustrate the key issues we can expand the first question from Table 2 – *Do bright disadvantaged children do worse (in terms of life-time outcomes) than their equally– talented, but socio-economic more advantaged, peers? Why*?

We have evidence that bright children from disadvantaged backgrounds do worse. But we do not know *why* disadvantaged children do worse. In particular, *at which life- stage(s) does poverty hinder bright children's chances*?

One view is that disadvantaged children's life-chances are largely determined in the early years of life.[xii] But many children may overcome early disadvantages and fall behind later. Some may even keep up with their peers from more advantaged social backgrounds until the reach employment and then fall behind as they are less able to access good jobs or progress in their careers. Information on this is limited.

Informed answers to social mobility and child poverty questions – and so intelligent public policy solutions – require sharing a lot of data and linking multiple databases.

Table 2: Questions to which greater access to administrative data would help improve answers**

- Do bright children from disadvantaged backgrounds do worse (in terms of life-time outcomes) than their equally–talented, but socio-economic more advantaged, peers? Why?

- Did the Educational Maintenance Allowance (EMA) have long-term benefits?

- Is it better to study *any* subject at an elite institution than a seemingly high-return subject (e.g., Law) at a less-prestigious university?

- Has the increased investment in early years' services over the last 15 years paid dividends in terms of improved social mobility?

- Has recent school improvement in London fed through into improved higher education and occupational outcomes?

- Did the recipients of Literacy and Numeracy Hour experience long-term benefits?

- Is the Pupil Premium boosting university applications?

- Are you less likely to be admitted to a Russell Group university if you are (i) poor, or (ii) from an ethnic minority? If so, why?

- How much do your peers at school affect your life-chances?

** Current knowledge is better where issues relate to social background and education. But this is precisely *because* some bodies (principally the Department for Education and Department for Business, Innovation and Skills) share data and permit data-linking.

What are the barriers to greater data sharing?

A significant challenge to data sharing is legal complexity and how this legal complexity interacts with institutional factors.

Potential *technical* problems around data sharing are not addressed in this report. For our purposes technical problems consist in the degree to which any data that is shared, in particular through the linking of different databases, is not of sufficient quality to be usable for research or public policy. For example, problems can arise where researchers cannot be confident that data from different datasets can be used to reliably match the correct individuals.

This report ignores technical problems because, while they are real issues, they are not barriers to *data sharing*. Once any technical problems are addressed the barriers to data sharing identified below will remain.

Legal issues

There are a number of legal issues which act as barriers to greater data sharing. Some of these relate to legitimate privacy concerns but many could be addressed through an improved culture of understanding of data-related legislation.

The Data Protection Act 1998 regulates the use and sharing of personal data. If data cannot be used to identify an individual it is not personal data. One way to remove the personal identifiers from data is to 'de-identify' it. De-identified data is data

> 'from which personal identifying information (names, addresses, exact date of birth, national Insurance number, national health service number, tax reference number, etc.) has been removed.'[xiii]

De-identifying data does not address all problems. It is impossible to reduce to zero the possibility that somebody can identify an individual from de-identified data. This is most obviously the case where datasets are linked. If a researcher has de-identified National Pupil Database data it may be practically impossible to identify specific individuals. But if her data is made richer – perhaps by linking it to UCAS, HESA, and HMRC data – it becomes much easier to identify specific individuals. The richer the data, the more people code 15T47K6 *cannot* be.

Privacy concerns are an extremely important constraint on data sharing but as the UK Statistics Authority notes

> 'The answer is not to lock up the administrative records in the organisation which first produced them so tightly that they cannot subsequently be used …. Rather, it is a matter of ensuring that statistical offices have all the right safeguards in place, and steadily building public confidence in their ability to handle confidential records.'[xiv]

As the Open Data White Paper notes, improving data sharing is not about 'compromising privacy and confidentiality; it is about using new technologies and techniques to analyse and access data safely and securely.'[xv]

A second legal issue is statutory prohibitions about the basis on which institutions can share data. HMRC, for instance, is permitted to share data only where doing so can be expected to aid it in carrying out its functions (i.e. where sharing data will facilitate the collection of tax revenue).

For many other bodies the statutory rules are not as prohibitive. But while most statute law creates 'gateways' for data sharing, it does not give institutions straightforward permissions to share data. Often it is unclear what the law is and what it permits. This necessarily contributes to risk-averse behaviour and unwillingness to share data. The Law Commission reports that gateways to share data

> '…are spread across a number of pieces of legislation, making them difficult to find and to interpret. For example, there are over 60 statutory gateways permitting the Department for Work and Pensions to disclose information to others, and far more provisions governing the onward sharing and use of information disclosed by the Department. These gateways are found in over 20 separate pieces of legislation.'[xvi]

Given this complexity it is understandable that there is

> 'a widespread lack of knowledge and understanding on the part of staff of public bodies of the circumstances in which information may be disclosed, to whom and for what purposes.'[xvii]

Summary

- *There are legitimate prohibitions on sharing personal data.*
- *It is impossible to completely 'de-identify' data.*
- *Some laws only permit data sharing for certain reasons.*
- *Most of the law around data is complex – often institutions are unclear about what they are permitted to do.*

Institutional and cultural factors

Ministers and government departments have priorities other than data sharing and sharing data is time- and resource-consuming.

In addition, most forces militate against permissive data sharing and linking. No civil servant was ever sacked for refusing to share data with another government department or researcher but some may have been disciplined for losing control over data. As it can be difficult to establish whether a particular individual can be trusted not to misuse data, sharing data can seem extremely risky.

A third problem is a misplaced sense of ownership. Some bodies can feel that it is their responsibility alone to generate findings from the data they control. They can see requests for information as interference to be ignored if possible. One illustration of this, reported by the Institute for Government, is that:

> 'A Downing Street adviser described the need for 'Sherlock Holmes-like skills' to extract information from departments and the Treasury – with a refusal in some cases to share even basic data on which policy is made.'[xviii]

The final, and most significant, issue is how the problems above interact with the legal complexity around data sharing. If incentives militate against data sharing, facilitating data requests is time-consuming, and it is unclear whether a department is even legally allowed to share data, one of two things will happen: data requests will be refused or only acted upon after a long period of time.

Summary

- *Facilitating data requests is time- and resource-consuming.*
- *At the level of individual civil servants, most forces militate against permissive data sharing.*
- *There can be a defensiveness about sharing data.*
- *All of these factors interact negatively with the legal complexity identified above.*

The proposed European Union Data Protection Regulation

In January 2012 the European Commission published a draft Data Protection Regulation. This regulation has been agreed by the European Parliament. As a *regulation* rather than a *directive,* the final version of the Regulation will apply directly to the UK.

The Commission's draft of the Regulation contained some exemptions for research. But European Parliament amendments narrowed these exemptions significantly. The 'use of personal data in research without specific consent would be prohibited or become impossible in practice.'[xix]

As all de-identified data starts as personal data, the European Parliament's amendment may apply to attempts to share more de-identified data for the purposes of research and public policy.

The UK research community believes that if made effective in its current form, the EU Data Protection Regulation would set-back the UK's ability to learn, assess and develop policies that might effectively improve social mobility and reduce child poverty.

A February 10 2015 joint press release by the Royal Statistical Society, the Joseph Rowntree Foundation and the Resolution Foundation illustrates the general concerns; it notes that

> 'If the [European parliament's] amendments are passed, it could put an end to many large-scale studies into social conditions and poverty, with millions of pounds of funding put at risk as research projects become too bureaucratic to continue.'[xx]

A case study: UCAS

UCAS has been identified by multiple sources as creating distinctive data sharing challenges. UCAS runs the applications process for UK higher education; it controls very important administrative data about the transition from school to university.

UCAS publishes aggregate-level statistics on university applications and admissions. But it has stopped giving researchers access to anonymised individual-level data. The social mobility implications of this are illustrated by research carried out by Durham University's Vikki Boliver.[xxi]

Using UCAS data, Boliver found that ethnic minority and state school applicants to Russell Group universities have to significantly out-perform their respective white and privately educated peers before they are as likely to be offered places.[xxii]

UCAS has disputed some of Boliver's claims, suggesting that much of the observed ethnic "bias" can be explained by ethnic minority students being more likely to apply to more competitive courses or with weaker predicted grades.[xxiii] However, as UCAS has not published this research or the data underpinning it, it is not open to scrutiny.

UCAS data would also be invaluable to answering a number of other important policy questions: for example, how much of the "disadvantage gap" in entry to elite universities after accounting for A-level grades[xxiv] can be explained by A-level subject choice and how competitive the degree courses they apply to study on are.[xxv]

University attendance is an important social mobility lever so getting a deeper understanding of the problems around access to elite universities is imperative. But public policy should only be based on research that can be independently tested and assessed. Obtaining a better understanding of social mobility in the UK requires that researchers and policy-makers have access to UCAS-controlled data.

UCAS must ensure that it satisfies its obligations under the 1998 Data Protection Act. UCAS also sees itself as having broader moral obligations to the individuals whose data it holds. Clearly, there is an important need to ensure that data sharing is safe. The Administrative Data Research Network is developing a system that will address this problem (see below).

Another potential blockage to data sharing is connected to how UCAS is funded. UCAS receives no direct public funding; it is financed by student application fees, institutional membership fees and revenue from its commercial activities.

UCAS Media – the commercial arm of UCAS – provides data intelligence services to educational institutions and the private sector; it also provides marketing services connecting businesses to university applicants through, for example, facilitating targeted emails to students who have just received their A-level results.[xxvi]

The profits from UCAS Media subsidise UCAS's charitable activities and the fees UCAS charges applicants and members.[xxvii] In all, UCAS Media provides about one third of UCAS's income.[xxviii] The need to protect UCAS Media's work, and the commercial value of UCAS data, is an important reason why UCAS can be reluctant to share data.[xxix]

Safely sharing UCAS-controlled data with researchers and policy-makers must become the norm. A way must be found to ensure that this is compatible with securing UCAS's long-term financial viability.

UCAS have said that they are committed to finding ways to make data available for research purposes and it is an issue that is under active consideration by the UCAS Board so hopefully rapid progress can be made.

What work is being done on the enable greater data sharing?

The problems identified above are widely recognised and there is ongoing action to address them. The three main actors in this area are the Cabinet Office, the Social Mobility Transparency Board, and the Administrative Data Research Network.

The Cabinet Office's Draft Data Sharing Bill

The Cabinet Office has worked on a draft Bill that includes a research and statistics section. The aim of the research and statistics section is to enable public bodies, except NHS bodies 'to engage, for the purposes of research, in the process of linking two or more datasets from two or more data controllers in a de-identified and therefore privacy enhancing way using a Trusted Third Party Sharing system which has been accredited under the legislation.'[xxx, xxxi]

The Trusted Third Party Sharing system (TTPS) is being developed by the Administrative Data Network (ADRN) – see below. The Data Sharing Bill will not be introduced in this parliament.

In total, the Data Sharing Bill has three sections: in addition to the research and statistics section, there is a fraud, error and debt section, and a tailored public services section. The fraud, error and debt section and the tailored public services section require sharing *personal data* – the research and statistics section does not.

As such, the fraud, error and debt section and the tailored public services section of the Data Sharing Bill are more likely to attract criticism. This may be based on the fear of data loss such as the 2007 HMRC loss of child benefit data.[xxxii]

The research and statistics section is likely to be the least controversial section of the Bill. This is supported by polling evidence:

> 'The ESRC and ONS commissioned research from Ipsos Mori published in 2014. The findings suggest that the public would be broadly happy with administrative data linking for research projects provided (i) those projects have social value, broadly defined (ii) data is de-identified, (iii) data is kept secure, and (iv) businesses are not able to access the data for profit.'[xxxiii]

In light of this, and because of the potential public benefits of greater sharing of de-identified administrative data, there is a strong argument for a separate Research and Statistics Data Sharing Bill**.**

The Social Mobility Transparency Board

The Government's social mobility strategy created the Social Mobility Transparency Board (SMTB). This strategic body meets twice a year; ultimately it reports to the Deputy Prime Minister and it is chaired by the Minister for Universities and Science. The SMTB aims to improve the government's use of data, especially in relation to individuals' progression through education and into the labour market.

The SMTB has had some successes. This includes removing the barriers to linking university student data to earnings data from HMRC to facilitate a project on the influence of university choice on graduate earnings.[xxxiv] The SMTB is also largely responsible for the Education Evaluation section of the Small Business, Enterprise and Employment Bill.[xxxv]

This part of the Bill 'will introduce measures to provide new and improved information on learning outcomes by tracking students through education into the labour market'[xxxvi]. It will also facilitate the 'sharing, at student level, [of] information on the destinations of former students with colleges in England and Wales.'[xxxvii]

Greater data sharing will 'inform Government about which qualifications and courses lead to sustained employment and higher incomes, and the link between family income and education outcomes'.[xxxviii] This information is essential in the design and assessment of policies that aim to improve social mobility.

The SMTB could be an important body in continuing to drive the social mobility agenda. It needs sustained ministerial leadership and involvement. Greater ministerial direction will help overcome institutional obstacles and enable government departments to work towards greater data sharing at a quicker pace.

The Administrative Data Research Network (ADRN)

An Economic and Social Research Council (ESRC) initiative led to a report by an Administrative Data Taskforce. The report - 'Improving Access for Research and Policy' – was published in December 2012.[xxxix] Its recommendations included the following:

1. 'An Administrative Data Research Centre (ADRC) should be established in each of the four countries in the UK.'[xl]

2. 'Legislation should be enacted to facilitate research access to administrative data and to allow data linkage between departments to take place more efficiently.'[xli]

ADRCs have now been created; together they form the Administrative Data Research Network (ADRN).

The central purpose of the ADRN is to 'give trusted researchers access to linked, de-identified administrative data in a secure environment.'[xlii] The ADRN has received £34 million of funding from the ESRC. The funding lasts until 2017.

The ADRN's proposal for safe data sharing

The ADRN has developed a system of safe data sharing. It enables researchers to obtain access to data without them seeing information through which they might identify individuals.

The proposal involves four agents and separating four different data elements. The four agents are an ADRC; an accredited researcher, a data controller, and a Trusted Third Party (TTP).

> 'A TTP is a trusted organisation with secure facilities for linking data, and will usually be independent of the organisations which hold data and of the Administrative Data Research Network.'[xliii]

The four different data elements are:

A. Information that identifies individuals, e.g., name or address;
B. Data about the individual, e.g., tax record;
C. A unique reference given to an individual, e.g., X1 = John Smith;
D. An Index Key – this connects the relevant unique references (C) from different datasets but does *not* include any data (B) or identifying information (A);

The ADRN's data sharing system ensures that none of the four agents see all of the data elements. This significantly reduces the privacy risk to individuals. Second, researchers can only access data in secure facilities; they cannot take data away with them. Third, all researchers who want access to data must be accredited by the ADRN. Each researcher must be judged to be a ''fit and proper' person capable of carrying out the research'.[xliv] Any project on which they work must also be approved by the ADRN. Projects are approved on the basis of 'feasibility, viability, ethics and public benefit'.[xlv] Finally, the ADRN can impose sanctions for data misuse. Beyond ordinary legal penalties, sanctions include the removal of funding from individuals or 'their entire institution'.[xlvi]

The ADRN is 'accountable to a governing board chaired by the UK Statistics Authority, which reports to Parliament'; and it is 'overseen in its day-to-day management by the ESRC, which reports to the Department for Business, Innovation and Skills.'[xlvii] The ADRN has no commercial purpose; it does not 'deal with commercial or private sector requests for access to administrative data'.[xlviii]

Given the ADRN's remit and its accountability mechanisms, it would make sense for it to be one of the main bodies through which greater data sharing is co-ordinated.

How can more progress be made? Recommendations

This report makes two kinds of recommendation. First, is a statement of three principles that should guide government's approach to data sharing. Second, there are 19 discrete policy recommendations.

The policy recommendations are very important but, in future years, new policy challenges will arise. When they do, an understanding and appeal to the three principles should guide government responses.

The Three Administrative Data Sharing Principles

I.   Where there is a clear public benefit, and where there is a system in which data can be shared safely, data should be shared.

II. Where legislative changes are required to realise principle one, they should be made.

III. Given the potential benefits of data sharing, the presumption should be that all bodies that control important administrative data should share their data. Bodies should expect to justify and be held accountable for any refusals to share data.

Quantifying the public benefits

There is a compelling argument that sharing more administrative data would lead to better public policy outcomes, both through reduced costs and more effective policy interventions. However, there is an absence of high-quality quantitative research on this subject.

This skews the debate against sharing data. While the potential privacy *costs* of greater data sharing are clear, the potential costs of *not* sharing more data are not.

Data sharing is an issue on which political leadership is necessary. But without reliable figures, it can be hard for politicians to take a lead and make the case for action.

1. *To address these problems, the UK research councils should fund research that quantifies the net public benefits of greater data sharing. To increase public confidence in the findings, this research should be peer-reviewed.*

   As part of this exercise all controllers of administrative data should provide details of the costs and benefits **to them** *of greater sharing. Where relevant, this should include the revenue they obtain from the sale of data they control, and the costs they would face from facilitating greater data sharing.*

Legislative changes

The Cabinet Office has worked on a Data Sharing Bill. Its purpose is to increase data sharing. In its current form, the Bill has three sections: a research and statistics section, a fraud, error and debt section, and a tailored public services section.

The fraud, error and debt section and the tailored public services section would require sharing *personal* data. The research and statistics section would only require sharing *de-identified* data. The research and statistics section of any Data Sharing Bill is likely to be the least controversial section.

It would be regrettable if the research and statistics section of any Data Sharing Bill did not become law because of objections to other sections of the Bill. No Data Sharing bill will be brought before this Parliament.

2. *In light of this, and because of the potential benefits of greater sharing of de-identified data, the current Data Sharing Bill should be abandoned.*

   When Parliament re-convenes after the May 2015 general election, the Cabinet Office should begin work on separate Data Sharing Bills. There should be a stand-alone Research and Statistics Data Sharing Bill. This Bill should be deemed a priority and brought before parliament before the end of this year.

   At the beginning of the next Parliament, the Cabinet Office and the Social Mobility Transparency Board should meet and make a collective response to this recommendation.

Political leadership

To ensure government departments share more data, the culture within government must change. This requires leadership from Ministers and senior civil servants.

3. *The Social Mobility Transparency Board has been a successful innovation. It should be retained and made responsible for leading and co-ordinating the creation of an effective and safe culture of data sharing across government.*

   To help it with this work, the Social Mobility Transparency Board should establish a formal working relationship with the Administrative Data Research Network's Board.

4. *The Social Mobility Transparency Board should ensure that, by September 2015, all Ministers review their department's data sharing policies. The results of each review should be published. Where there is a need for change, each minister should establish data sharing policies within his/her department on the basis of the Three Administrative Data Sharing Principles stated above.*

   Each department's published results should include a list of the kinds of data it controls, whether that data is available to accredited researchers; and, if not, when it will be available to them.

5. To aid transparency the government should publish data requests and give a breakdown of acceptances and refusals along with reasons given for refusals. This will act as a very useful data-gathering exercise.

Administrative Data Research Network (ADRN)

The Administrative Data Research Network is developing a system through which it is safe to share de-identified administrative data. This 'Trusted Third Party System' (TTPS) has great potential to improve data sharing. However, funding for the ADRN lasts only until 2017.

6. Subject to a value-for-money evaluation of its work, funding for the ADRN should be made permanent. Funding should come from the ESRC alone or in conjunction with other research funding bodies.

7. The ADRN should be given a formal role to engage with and address the general public's legitimate worries about data sharing. This should include production and promotion of data sharing 'myth-busting' documents aimed at public bodies and the general public.

Reducing time and resource costs

One problem government departments and public bodies face in facilitating data requests is in establishing whether they are legally permitted to share the requested data. It would be helpful if, where necessary, data requests came with legal advice attached. This is not easy to achieve. However, the ADRN would be the ideal body to facilitate this advice.

8. Therefore, where there is legal uncertainty, it is recommended that the ADRN Board considers the feasibility and costs of obtaining legal advice for the data-requests it facilitates through its TTPS.

9. A separate budget should be created to fund the satisfaction of data requests. Compensation to departments and other public bodies that control administrative data should operate on a standardized time-cost basis. The Social Mobility Transparency Board and the ADRN Board should meet to organize the calculation of a standardized time-cost for facilitating data-requests.

There is no reason why recommendation nine should increase *net* government expenditure. In light of the potential savings from greater data sharing, medium- and long-term savings should be expected.

One problem data requesters face in accessing data is a lack of clarity about who, within in the relevant body or department, has data control responsibilities. Perhaps surprisingly, this problem is as common *between* government departments as it is between government departments and other public bodies or individuals.

10. Each government department, or relevant public body, should establish the role of Head Data Sharing Officer. The role-holder's name and personal email

address should be made available to all ADRN-accredited researchers as well as to individuals in government departments and other public bodies. The Head Data Sharing Officer's brief is to enable and facilitate data sharing.

It is not envisaged that this would be a full time role; it is a responsibility that can be added to the existing role of a senior official.

11. *The government should also establish the role of Chief Data Sharing Officer to whom all Head Data Sharing Officers are accountable. The Chief Data Sharing Officer's name and personal email address should also be available to all ADRN-accredited researchers, and to individuals in government departments and public bodies. The Chief Data Sharing Officer's brief is to ensure that all Head Data Sharing Officers are enabling and facilitating data sharing. The Chief Data Sharing Officer should also drive good practice in the civil service and the wider public sector.*

It is the norm for individual government departments to determine their own data sharing policies. This leads to unnecessarily different, and inconsistent, approaches to data sharing. Excepting special circumstances, there is no need for different departments to approach data sharing differently.

12. *The Social Mobility Transparency Board and the Administrative Data Research Network should be tasked with the role of establishing national standards for data sharing. Excepting special cases, these standards should be adopted by all government departments and other bodies that control administrative data.*

Non-government bodies with public responsibilities

Important administrative data is held by bodies other than government departments. The autonomy of these bodies should be respected. But the data sharing expectations of non-government bodies that control administrative data should be the same as the data sharing expectations of government departments. Such bodies should expect to justify and be held accountable for any refusals to share data.

UCAS

UCAS controls important data about the transition from school to university. More of UCAS's data should be shared with public policy researchers and other public bodies.

13. *UCAS should provide details to the SMTB of the costs it would face if it shared de-identified data on request through the ADRN's Trusted Third Party System. This figure should be included in the calculation to quantify the net public benefits of greater data sharing.*

14. *The relevant UCAS Executives should give details to the SMTB of any of UCAS's data-security concerns about sharing data through the ADRN's Trusted Third Party System.*

15. *The ADRN Board should respond to UCAS's concerns and, where necessary, alter its data sharing system.*

16. *Given the importance of UCAS data to social mobility and child poverty questions, consideration should be given to inviting UCAS to become a core member of the SMTB.*

17. *The SMTB should ensure that recommendations 13-16 are acted upon satisfactorily by 31 December 2015.*

18. *In addition, the SMTB should ensure that, from 1 January 2016, data sharing of UCAS data through the ADRN TTPS is the norm.*

European Union Data Protection Regulation

In January 2012 the European Commission published a draft Data Protection Regulation. The Commission's draft of the Regulation contained exemptions for research. But European Parliament amendments narrowed these exemptions significantly.

If made effective in its current form, the EU Data Protection Regulation would seriously weaken the ability to legally share de-identified administrative data.

19. *Research Council UK (RCUK) should approach other European research bodies and work with them to amend the European Union Data Protection Regulation so that it does not inadvertently set back social mobility and other important research. RCUK should co-ordinates its efforts with the Social Mobility Transparency Board and other relevant bodies.*

Appendix 1 – Data sources for researchers and policymakers

Source	Problem/limitation
Cohort studies	Cohort studies are expensive and suffer from rates of attrition over time. For instance, in a recent Annual Social and Economic Supplement of the US's Current Population Survey non-response rates for some issues were over 30%.[xlix]
Samples from UK administrative databases	A researcher may want to look at a particular ethnic minority which may constitute only 5% of the whole population. If she is given a 10% sample she may have too little data to produce meaningful findings. For instance, in breaking down a sample of data from the UK's Annual Population Survey, it was found that the sample included only 19 Pakistani students who were doing Single Honours Medicine.[l] Meaningful conclusions cannot be drawn from such a small sample. The UK Statistics Authority notes that, relative to sample surveys, the benefits of administrative data include 'reduced burden on respondents; low cost production processes; better quality statistics for small geographic areas; more frequent updating, usually; and the capacity to link information from different sources to produce richer data.'(Monitoring Brief (16 March 2012) - *Creating official statistics from administrative data*, p.1)
Research findings from other countries	Research may not generalize from one country to another. As noted on one Swedish study on peer effects: 'it must be emphasized that … [the findings] refer to the Swedish case and may not be generalizable to other cases' 'Choice by Contrast in Swedish Schools: How Peers' Achievement affects educational Choice', Jansen and Mood in *Social Forces Vol. 87 No 2* Dec 2008: 741-765, p.760. Equally, Raj Chetty notes that 'many questions of central importance for US policy making cannot be tackled using evidence from other countries. [For example, the] decentralized US labor market is quite different from the European labor market.'[li] As the UK labour market is also very different from many European countries' labour markets, the point applies to the UK too.

Appendix 2: definitions[lii]

Administrative Data 'information collected primarily for administrative (not research) purposes. This type of data is collected by government departments and other organisations for the purposes of registration, transaction and record keeping, usually during the delivery of a service.'

Data Linkage 'is the joining of two or more administrative or survey datasets using individual reference numbers/ identifiers or statistical methods such as probabilistic matching.'

Anonymised Data 'data have all personal identifiers removed and cannot be connected to the original person record.'

De-identified Administrative Data data for which 'personal identifying information (names, addresses, exact date of birth, national Insurance number, national health service number, tax reference number, etc.) has been removed.'

Pseudonymised Data 'data [that] cannot directly identify an individual as the personal data have been removed, but they include a unique identifier that enables the person's identity to be re-connected to the data by reference to separate databases containing the identifiers and identifiable data. The unique identifier allows datasets to be linked together, without knowing the identity of the person.'

Endnotes

[i] See Appendix 2 for definitions.

[ii] Other possible quantifiable benefits such as greater economic growth are ignored. For a report that looks at the potential economic benefits of better data-use see *Market Assessment of Public Sector Information* a report by Deloitte for BIS (May 2013).

[iii] https://www.gov.uk/government/policies/improving-the-transparency-and-accountability-of-government-and-its-services/supporting-pages/the-open-data-institute (accessed 23 February 2015).

[iv] See its tenth report 'Statistics and Open Data: Harvesting unused knowledge, empowering citizens and improving public services' (March 2014).

[v] 'Data Sharing between Public Bodies a Scoping Report' (Law Com No 351, July 2014).

[vi] http://www.statslife.org.uk/news/1789-society-launches-data-manifesto-for-general-election (accessed 23 February 2015).

[vii] Dr. Dean Machin teaches in the Philosophy Department at University College London (UCL). He has written this report as part of UCL's Policy Placement Programme. He is a former Leverhulme Fellow, chairs a UCL policy commission into transport and ethics, and blogs at www.deanmachin.com.

[viii] UK Statistics Authority Monitoring Brief (16 March 2012) *Creating official statistics from administrative data,* p.5.

[ix] Letter of Chair of the UK Statistics Authority, Sir Michael Scholar (16 March 2012) See http://www.statisticsauthority.gov.uk/reports---correspondence/correspondence/letter-from-sir-michael-scholar-to-rt-hon-francis-maude---administrative-data---16032012.pdf (accessed 3 March 2015).

[x] Chetty, Hendren, Kline, and Saez 'Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States', National Bureau of Economic Research Working Paper No. 19843 (Jan 2014 – revised June 2014). See http://www.nber.org/papers/w19843 (accessed 23 February 2015)

[xi] See Appendix 1 for a fuller explanation of the relative merits of administrative data.

[xii] See, instance, 'Promoting Social Mobility' *The Boston Review* (September 2012).

[xiii] 'Improving Access for Research and Policy' (December 2012), p.41. See http://www.esrc.ac.uk/_images/ADT-Improving-Access-for-Research-and-Policy_tcm8-24462.pdf (accessed 24 February 2015).

[xiv] UK Statistics Authority Monitoring Brief (16 March 2012) *Creating official statistics from administrative data*, p.3

[xv] *Open Data – unleashing the potential* (Cmnd. 8353, June 2012), p.37.

[xvi] 'Data Sharing between Public Bodies a Scoping Report' (Law Com No 351), pp.26-7, §1.107

[xvii] Ibid.

[xviii] 'Centre Forward: Effective Support for the Prime Minister at the Centre of Government' (IfG, July 2014), p.17. See http://www.instituteforgovernment.org.uk/sites/default/files/publications/Centre%20Forward%20-%20Final.pdf (accessed 9 January 2015).

[xix] 'Protecting health and scientific research in the Data Protection Regulation (2012/0011(COD)) Position of non-commercial research organisations and academics' – July 2014, p.1. See http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/WTP055584.pdf (accessed January 5 2015).

[xx] 'New EU proposals threaten UK poverty research', Royal Statistical Society, 10 February 2015. See http://www.rss.org.uk/Images/PDF/about/press-releases/2015-02-10-eu-proposals-threaten-uk-poverty-research.pdf (accessed 23 February 2015).

[xxi] The data was supplied by UCAS *before* it decided to stop providing individual-level data to researchers.

xxii Boliver, V (2013) 'How fair is access to more prestigious UK Universities?' *British Journal of Sociology* 64(2): 344-364. For instance, "when those from state schools do apply to Russell Group universities they seem to need to be better qualified than their private school counterparts by as much as a B-grade A-level before they are as likely to receive offers of admission. Black and Pakistani/Bangladeshi applicants to Russell Group universities seem to be similarly disadvantaged in comparison with White applicants when it comes to the likelihood of being offered a place at a Russell Group university."(p.359).

xxiii "Small' Russell Group racial bias in admissions: Ucas' *Times Higher Education, 2 May 2013. Available at:* www.timeshighereducation.co.uk/news/small-russell-group-racial-bias-in-admissions-ucas/2003594.article (accessed 23 February 2015).

xxiv For example, Social Mobility and Child Poverty Commission, Higher Education: The Fair Access Challenge (2013) identified 3,700 "missing" students from state schools and 1,900 "missing" students from working-class backgrounds at Russell Group universities based on what would be expected if they accepted a socially representative sample of students with the right A-level grades to enter

xxv For example, see Russell Group, New Office for Fair Access Strategy (26 February 2015)

xxvi Emails can target students by 'the subject they have applied to study, where they are living, their age, the institutions they have applied to plus their predicted grades at A-level.' 'Clearing email and tweet rate 2014'. See https://www.ucasmedia.com/sites/default/files/Clearing%20Email%20and%20Tweet%20Rate%20Card%202014_1.pdf (accessed 28 January 2015).

xxvii 'Why UCAS Media?' https://www.ucasmedia.com/about/why-ucas-media (accessed 28 January 2015).

xxviii ICO Freedom of Information Act 2000 (FOIA) Decision notice Reference: FS50453565, paragraph 63. See https://ico.org.uk/media/action-weve-taken/decision-notices/2013/831630/fs_50453565.pdf (accessed 10 March 2015).

xxix In response to a freedom of information request, the risk of "prejudice [to] its … commercial interests"(ibid., paragraph 60)xxix has been cited by UCAS as a reason to withholding data (ibid., paragraphs 60-63).

xxx 'Part One – Draft De-identified Strand Paper' (8 Aug 2014 draft), p.4 at http://datasharing.org.uk/latest/ (accessed 10 Sept 2014).

xxxi The NHS is excluded because 'de-identified data from them can be made available within the Health and Social Care Information Centre, and the Department of Health does not believe it is appropriate for another or alternative method of disclosure is to be made available.'(ibid., p.9).

xxxii 'Data lost by Revenue and Customs' (21 Nov 2007) http://news.bbc.co.uk/1/hi/uk/7103911.stm (accessed 23 February 2015).

xxxiii Summary of civil society and public sector policy discussions on data use in government'' http://datasharing.org.uk/interim-findings/ (accessed 14 November 2014).

xxxiv See 'Estimating human capital of graduates: the influence of institution, gender, year of entry into higher education and subject' IFS (January 2014) See http://www.ifs.org.uk/docs/graduates_info.pdf (accessed 10 March 2015).

xxxv On 3 March 2015 the Bill will go to the Report Stage in the House of Lords. See http://services.parliament.uk/bills/2014-15/smallbusinessenterpriseandemployment.html (accessed 23 February 2015).

xxxvi 'Small Business, Enterprise and Employment' Bill, p.3. See https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/376998/bis-14-938-small-business-enterprise-and-employment-bill-summary.pdf (accessed 5 January 2015).

xxxvii Ibid., p.3 (accessed 5 January 2015).

xxxviii http://www.publications.parliament.uk/pa/bills/lbill/2014-2015/0057/en/15057en.pdf, p.13 (accessed 9 January 2015).

xxxix See http://www.esrc.ac.uk/_images/ADT-Improving-Access-for-Research-and-Policy_tcm8-24462.pdf. Accessed on multiple dates.

xl Ibid., p.iii.

xli Ibid., p.v.

xlii http://www.adrn.ac.uk/about/what-we-do (accessed 20 November 2014).

xliii http://www.adrn.ac.uk/admin-data/anonymous (accessed 24 October 2014).

xliv http://adrn.ac.uk/faq/about-the-researchers (accessed 24 February 2015).

xlv http://www.adrn.ac.uk/about/values (accessed 20 November 2014).

xlvi http://adrn.ac.uk/faq/privacy-and-security (accessed 26 January 2015).

xlvii http://www.adrn.ac.uk/about/values (accessed 20 November 2014).

xlviii http://www.adrn.ac.uk/about/values (accessed 20 November 2014).

xlix 'Expanding Access to Administrative Data for Research in the United States' by Card, Chetty, Feldstein, and Saez see http://eml.berkeley.edu/~saez/card-chetty-feldstein-saezNSF10dataaccess.pdf, p.2 (accessed January 7 2015).

l Jo Wathan (UK Data Service) 'Key Data and their Issues', presentation at the Royal Society *Diversity in STEMM: Challenges to using, manipulating and managing data* event, The Royal Academy of Engineering, 4 November 2014.

li 'Expanding Access to Administrative Data for Research in the United States' by Card, Chetty, Feldstein, and Saez see http://eml.berkeley.edu/~saez/card-chetty-feldstein-saezNSF10dataaccess.pdf, p.3 (accessed January 7 2015).

lii All definitions taken from 'Improving Access for Research and Policy' (December 2012), pp.41-2. See http://www.esrc.ac.uk/_images/ADT-Improving-Access-for-Research-and-Policy_tcm8-24462.pdf (accessed 23 February 2015).