

CABINET OFFICE INITIAL DISCUSSION DOCUMENT ON DATA SHARING POLICY FOR PUBLICATION

9 APRIL 2014

Introduction

People tend to assume that Government can share data between departments to complete simple tasks, and are surprised to learn that it cannot. Removing barriers to sharing or linking different datasets can help Government to design and implement evidence based policy, for example to tackle social mobility, assist economic growth and prevent crime.

We are acutely sensitive to the potential concerns of citizens and proposals need to be designed in a way that safeguards people's privacy. We are keen to undertake an open policy making approach to this work through bringing together relevant parts of Government with stakeholders who have an interest in the use of data for delivering better public services. We recognise that the views and opinions in relation to data sharing are many and diverse. Our ambition with this work is to listen to and understand the arguments advanced to help us develop proposals that will help deliver necessary changes and resultant improvements to public service delivery and the lives of citizens.

Our planned approach, beyond the initial open policy making period and assuming good progress can be made in agreeing policy proposals, is to subject this work to some form of scrutiny or wider public consultation. Any decision to introduce draft legislation into Parliament will be taken at a later date.

The focus of this work is data sharing across public bodies but does not involve the care.data initiative which is led by NHS England.

Three initial strands of focus

There are currently **three key strands to our thinking** - research & statistics, tailored public services for individuals, and fraud, error & debt.

Research and statistics

To improve our understanding of the UK's economy and society, and:

- Make it easier for the Office for National Statistics (ONS) to access data from public authorities, to enable it to better carry out its functions. This would help with the production of more accurate estimates of GDP to aid fiscal and monetary policy formulation, and reduce both the administrative burden on employers/businesses from surveys and the operational costs to government;
- Provide ONS with options for the future of the census, e.g. by using administrative data already collected by government supplemented by mandatory population surveys.

Sharing of de-identified data could provide new opportunities to:

- Help develop effective policies to support young people by identifying pathways to success, and barriers to social mobility by linking data on education, employment status and income.
- Improve energy efficiency and save citizens money by linking data on energy use with property data;
- Help deliver targeted crime prevention strategies.

Creating tailored public services for individuals

Innovative and tailored approaches to public service delivery are essential to addressing key social challenges, such as long term unemployment and preventing families spiralling into crisis. More effective use of data has significant potential to support this transformation and also deliver more efficient and cost-effective public services.

Examples of the potential benefits from more effective and efficient data sharing include:

- Data sharing between departments and local authorities to target energy efficiency measures and fuel poverty grants, reducing mortality rates and hospital admissions amongst vulnerable groups;
- Better identification of families requiring more assistance and targeting of services and support, reducing costs to government and delivering better outcomes for those most in need.

This could be in the form of a permissive but constrained power to share data between defined public agencies for specified purposes such as the delivery or targeting of public services for individuals from specified groups. The aim would be that individual whose data is shared would benefit through, for example, improved outcomes in health, education or employment.

Fraud, Error and Debt

The tax-payer is losing an estimated £37 billion to fraud, error and debt annually. Those committing fraud exploit the slowness of the system by changing tactics regularly. This leaves public authorities 'playing catch-up'.

A more holistic view of an individual's debt with Government can lead to better managed repayment, whilst relieving the pressure that mounting debt can place on those most at need.

The fraud, error and debt proposals could allow specified organisations to share any data for the purposes of the prevention, detection, investigation and pursuance of fraud, error and debt.

Safeguards

The aim is to design the proposals for each strand so as to safeguard privacy, taking a standard approach across the proposals as far as possible, while ensuring the unique needs of each policy area are preserved.

Current thinking

Research and Statistics:

ONS

Following the conclusion of an extensive programme of research and a three-month public consultation on “The census and future provision of population statistics in England and Wales,” on 27 March 2014 the National Statistician made a recommendation to the Board of the UK Statistics Authority. The National Statistician has recommended a predominantly online census in 2021 supplemented by further use of administrative and survey data.

ONS could be given powers to conduct mandatory population surveys. Regulations may then set out circumstances in which a survey could be conducted. The regulations could be similar to the powers under the Census Act 1920. Offences in relation to population surveys could be similar to offences under the Census Act 1920 in respect of the census.

ONS could also be helped in carrying out its functions by receiving more data held by other parts of Government. The Statistics and Registration Services Act 2007 could be amended to authorise the disclosure of information held by public authorities to ONS for ONS’ functions (which are defined in the SRSA). Information from HMRC, for example, could allow ONS to improve the quality and speed of estimates of GDP, which is a key measure for both fiscal and monetary policy formulation.

Trusted Third Party Proposals

In many cases, research on Government and public body data is limited to the analysis of single data sets which, if the researchers are external to Government, is almost always of anonymised or de-identified data. Consequently the possibility of undertaking deeper research using cross-linked but separate datasets is difficult or impossible. Bodies holding the data can be reluctant to undertake such shares and analysis because they lack the necessary powers to share information or are subject to a statutory bar. Those outside Government can find it difficult to access data in the first place and then cross-link two or more de-identified sets of data.

To assist in overcoming these issues, the report by the Administrative Data Taskforce *Improving Access for Research and Policy*¹ (pages 44 to 46 of which contains a useful Q&A) recommended a model of data sharing that allowed for such cross-linked research on de-identified data to take place, whilst at the same time maximising privacy protection for data subjects by restricting access to and the use of identity data to the absolute minimum required to cross-link the datasets. There are two variants of the model (the Trusted Third Party variant and the Firewall Single Centre variant), but the majority of structural elements are the same for both.

¹ http://www.esrc.ac.uk/_images/ADT-Improving-Access-for-Research-and-Policy_tcm8-24462.pdf

Some but not all public bodies are able to use either variant without further provision being necessary, but there are some public bodies that are currently either entirely or partially prohibited from using either variant to share data with other bodies for research purposes other than their own.

Future legislation could remove the initial constraints. Such shares would be dependent on the agreement of the data controllers involved and whether the circumstances and agreements comply with the other legal requirements, in particular the DPA and HRA. The variants are essentially a complicated method of data-sharing between A and B (or from public bodies to an ADRC). The added complexity is beneficial because it de-identifies the data and ensures that **no participant in the process** ever controls the complete set of payload and identity data. A diagram illustrating the data flows for the Trusted Third Party model is attached.

1. Proposals could provide all UK public bodies with a discretionary power to:
 - a. Disclose both personal data and other data to an accredited external data processor (who could be either a public or a private body but not one of the data owners who are the source of the information) for the purpose of indexing against another dataset (which has been provided on the same basis from another data owner). Data owners would be able to disclose whatever type of data is necessary in order for the indexer to be able to match records between the datasets. For example, in many cases an address will be a useful type of data to use as an identifier, but if the research focused on homeless people then it is unlikely to be useful; another data type might be used instead.
 - b. Share **de-identified** personal and other data with accredited ADRCs², or alternatively disclose de-identified personal and normal data to ADRCs (where they act as a data processor).
2. These powers and one or other of the Trusted Third Party sharing methods could provide the opportunity to make more data available for research purposes. Where two public bodies wish to conduct research that requires matched data from both bodies they could use the system and make both the ADRC and any researcher data processors on behalf of both of the public body data controllers jointly. Once analysed the joint database could either be delinked or destroyed. Non-personal data containing research products could be used and retained by both public bodies, providing this is in accordance with other legislation.

Such a power would be subject to the conditions which could include:

- the purpose of the share or disclosure is to enable processing for research or statistics purposes (as per s.33 of the DPA);

² The Economic and Social Research Council (ESRC) is, as a result of the December 2012 report of the Administrative Data Taskforce, in the process of establishing a number of new safe settings outside and independent of Government, in the form of safe havens called **Administrative Data Research Centres** (ADRCs) in each nation of the UK.

- access to the payload and identity data during the TTP share process is restricted to accredited persons.

Safeguards

Individuals who have access to the de-identified data could be required to be accredited, and also the projects for which de-identified data is sought. A register of all individuals and projects which have been accredited could be published.

The four Administrative Data Research Centres would be the initial bodies to be accredited by the UKSA as safe havens and repositories for the de-identified data.

The Data Sharing Code of Practice and Anonymisation Code of Practice, published by the ICO, would remain.

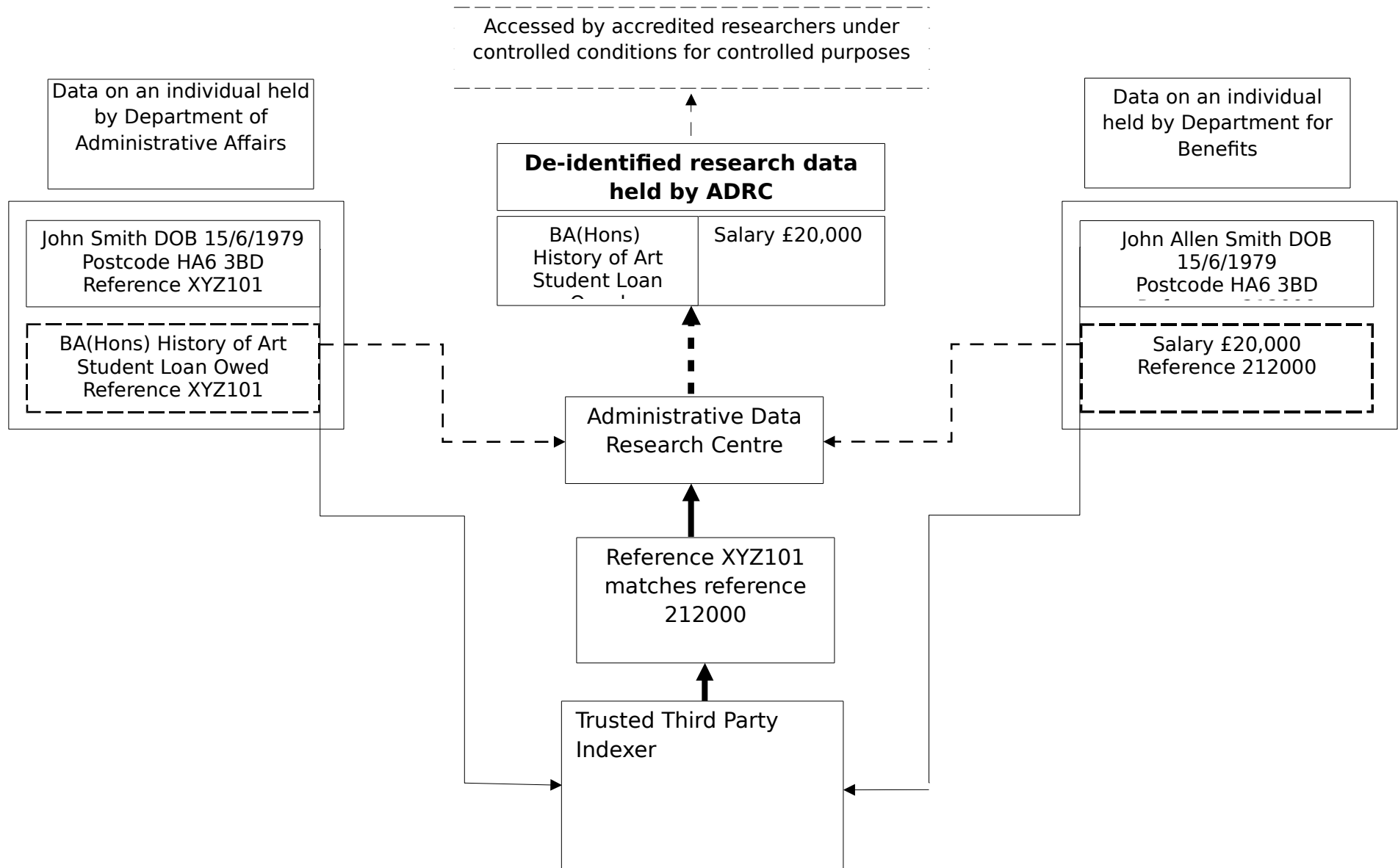
Potential Outcomes

The value that could be derived from easier linkage of de-identified data includes:

- addressing inequalities of access to public services and social mobility/ outcomes – by linking data on education, training, employment, unemployment, incomes and benefits;
- improved energy efficiency and building stock– by linking data on energy use with property data;
- crime prevention and improved community safety by linking data on (re)offending behaviour, incomes and benefits);
- researching causal pathways over the life course – linking data on education, employment, incomes and wealth;
- informing policies designed to tackle poverty – linking data on housing conditions, incomes and benefits.

Value can also be gained from linking Government data to other studies, including ongoing longitudinal and other surveys, which these powers could facilitate.

UNCLASSIFIEDUNCLASSIFIEDTrusted Third Party Model, showing how data from two different departments are de-identified (dotted line) and linked using their identifiers (solid line), ending as de-identified, linked data in an ADRC (dotted heavy line).



Tailored Public Services:

New ways of providing services are essential to improving quality and addressing cross cutting social challenges. More effective use of data is a tool with significant (and proven) potential to benefit individuals and society and is key to supporting the transformation of public service provision. The work **to better tailor services to individuals could include** the creation of powers to allow organisations to share data around specific groups of citizens who use multiple public services for the purposes of improving their health, education and employment.

Current thinking

Provide an ability to share data between defined public agencies for the purposes of improving the delivery or targeting of public services for individuals from specified groups, where the individual whose data is shared benefits from the share through improved outcomes in health, education or employment.

Public Agencies

We want to identify the best solutions to facilitate data sharing amongst the public, private and third sector agencies which provide public services in a manner that will benefit the citizen while protecting their information. There are a number of options to explore, for example: including all bodies providing public services; excluding all non public agencies that are providing public services; listing non public agencies to be included by named type of body; listing non public agencies to be included by type of relationship; including all public service providers but limiting the data share to one way (upwards) etc. Key bodies may include:

- Government departments
- Local authorities
- Local emergency services
- Police
- Schools

Potential Groups

Specific groups would be identified because of a particular data sharing need which, if met, would improve the delivery of public services to them and improve their quality of life. We should consider how we could future proof any legislation introduced. Groups could include:

- Households with multiple disadvantages
- Vulnerable elderly
- Ex-offenders and current offenders
- Gang members

- Long term unemployed
- NEETs/ 19-24 year olds who are unemployed
- Carers

Rationale

A legal framework could enable data shares around particular groups of service users. It should be proportionate to the benefits.

Data shares that meet the criteria could take place without further legislation (although still subject to the protections in the DPA). A non-legislative process of approval could be considered (see "Safeguards" below).

Examples of the potential benefits from more effective and efficient data sharing include:

- Data sharing between departments and local authorities to target energy efficiency measures and fuel poverty grants, reducing mortality rates and hospital admissions amongst vulnerable groups;
- Better identification of families requiring more assistance and targeting of services and support, reducing costs to government and delivering better outcomes for those most in need.

Safeguards

A key element of the proposal would be safeguards to ensure that personal information is protected. Options include, for example:

- transparency of data shares so that the public are fully informed of the process
- a robust offence of misuse of data
- exploring the best technical processes for storage and disposal of any data held
- how best to ensure that consent is used where appropriate
- how best to ensure that anonymous data is used where possible
- determining how the specific process of applying to share data will work

Fraud, error and debt:

The Need:

1. A recent estimate is that there are 92 gateways for sharing data on debt and over 230 on Fraud. This has led to an inconsistent patchwork quilt of legislation that is difficult and time-consuming to navigate. Where powers don't exist to share data, traditionally new legislative relationships have been created, constantly adding to this confusing area of the law.

2. This also provides an opportunity to look at ways that data can be used to reduce administrative and system error by comparing data sets to ensure that public authorities aren't wasting time and tax-payer's money in delivering services that aren't needed. For example, identifying deceased recipients of locally delivered benefits (such as freedom passes).

Current thinking

3. The **Fraud, Error and Debt** proposals could create **permissive gateways** to allow specified, listed organisations to share any data for the purposes of the **prevention, detection, investigation and pursuance of fraud, error and debt**, constrained by limiting the **organisations** that can share data and the **purposes** that the data can be shared for, alongside the use of appropriate safeguards.
4. A Code of Practice could be created, and options could include ensuring the publication of privacy impact assessments.

Operation of proposals

- a. Any public authority or organisation providing services of a public nature on behalf of a public organisation that isn't specified at the outset could apply to join the lists of those who can share data for these purposes. The addition would be made by secondary legislation.
5. Once in the list, an organisation could request data from another organisation on the list for the purposes of the gateway (Fraud, Error or Debt). The data holding organisation would still have to consider its duties under the DPA.
6. The Minister could have the power to remove an organisation if there is evidence of non-compliance with the Code of Practice by the organisation or those providing services to it.

Rationale

A proposal should aim to strike a balance between regard for privacy and the need to improve the effectiveness and efficiency of public services for the wider public good.

Implementation examples

7. In Fraud, this power could enable things like:
 - a. The identification of NHS Bursary Fraud;
 - b. The use of housing benefit data to detect and investigate tenancy fraud;
 - c. The identification and investigation of Land Registry Fraud (changes of address).
8. In Error, this power could enable sharing to resolve issues such as:
 - a. Validating that a person in receipt of a freedom pass is not deceased;
 - b. Validating the student status of those in receipt of a Council Tax rebate;

9. For Debt, this power could enable consolidation of an individual's debt to Government, allowing more manageable repayment plans and more efficient collection of debts on the part of public authorities.