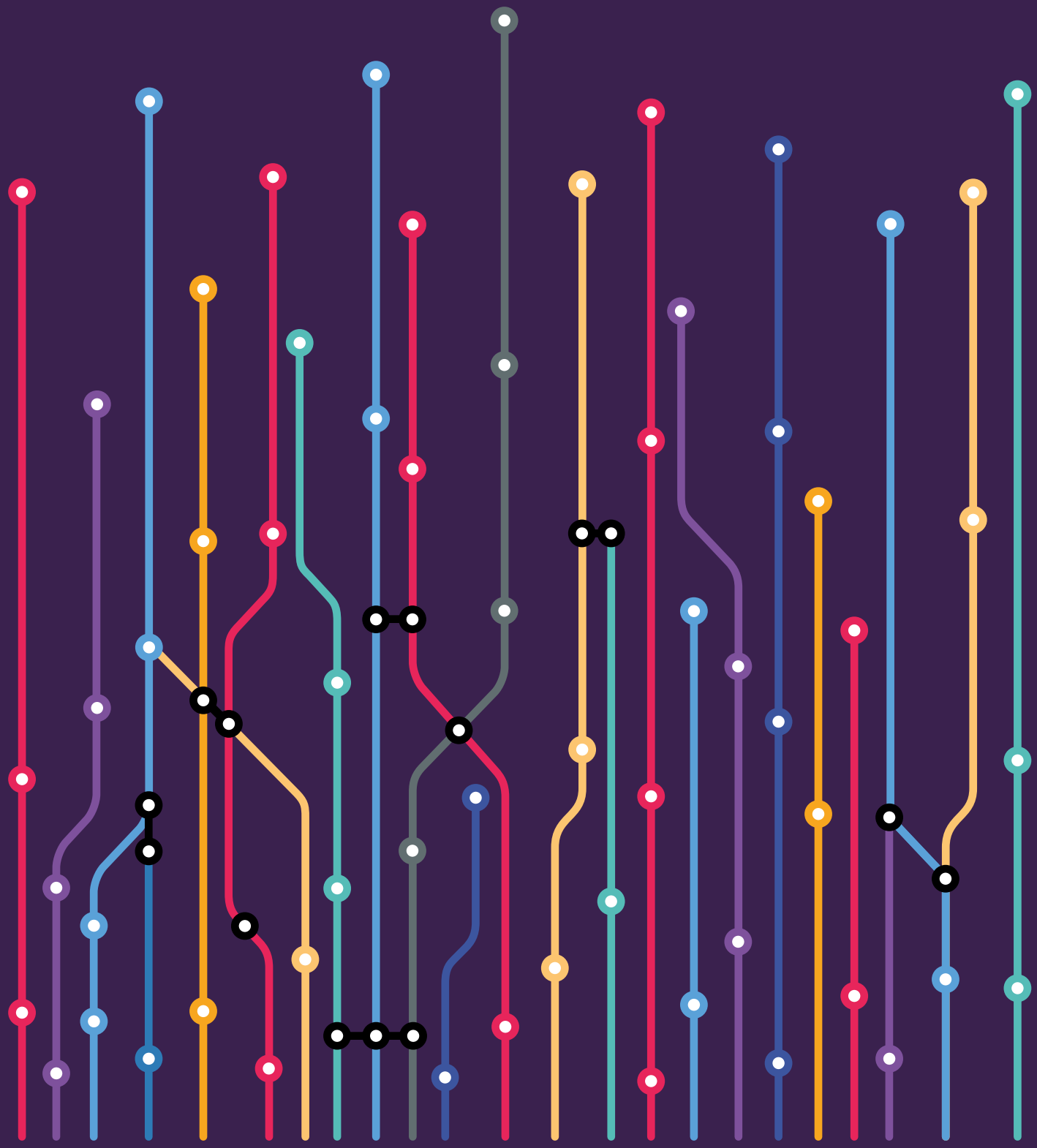
The UK Administrative Data Research Network:
Improving Access for Research and Policy

Report from the Administrative Data Taskforce
December 2012



The UK Administrative Data Research Network:

Improving Access for Research and Policy

Report from the Administrative Data Taskforce – December 2012

The UK has the opportunity to be a world leader in research using de-identified administrative data, routinely collected by government departments, agencies and other statutory bodies. Such data, made accessible for research in ways that prevent the identification of individuals, will provide a robust UK-wide evidence base to inform research, thereby guiding the development, implementation and evaluation of policy. Meeting this aim requires improvements in procedures for access to and linking between such data. This entails not just the development of safe, secure and efficient systems for linking, managing and analysing administrative data, founded on secure technologies, but on further building of trust between data providers, researchers and all other interested parties. The new system must adopt the highest international standards of governance, professional practice and public engagement. New legislation needs to be introduced to enable efficient data linkage, and government and the relevant funding agencies need to resource the new system to ensure its integrity, sustainability and utility.

The Administrative Data Taskforce was formed in December 2011 with the aim of improving access to and linkage between government administrative data for research and policy purposes. This initiative was led by the Economic and Social Research Council in collaboration with the Medical Research Council and the Wellcome Trust.

Contents

Foreword

The Department for Business, Innovation and Skills' *Innovation and Research Strategy for Growth* published in December 2011 committed to a wide range of policies for stimulating innovation and growth and improving the competitiveness of UK research.

In addition to confirming ring-fenced science and research funding of £4.6 billion per annum, the commercialisation of advanced technologies and new tax incentives for research and development, the government underlined the importance of improving access to administrative data and publicly funded research. *The Open Data* White Paper published by the Cabinet Office earlier this year reinforced the specific challenge of ensuring effective use of national data collections for research and policy development.

This report meets that challenge head on. Drawing on international best practice and the progress made in the UK so far, it shows how improved access to, and linkage between, datasets will ensure that we can tackle some of the major issues facing society and the economy in new and innovative ways. By unlocking the research potential of these data, we will improve our knowledge and understanding of the action required to tackle a wide range of social, environmental, health and security issues – confirming the leading international reputation of UK universities and research institutes and promoting new approaches to the development, implementation and evaluation of policy across government.

The report is clear that progress in this area must be founded on **trust**. It argues for safe and secure systems for managing and linking data; the highest international standards of governance, professional practice and public engagement; primary legislation to establish a generic gateway for research and statistical purposes that enables efficient access to, and linkage between, data held in different parts of the public realm; and the resources required to ensure the integrity, professionalism and coherence of robust UK-wide systems and processes for delivering on this promise.

The *Administrative Data Taskforce* was able to draw on the expertise of many government departments, a range of research funders, the Office for National Statistics, representatives from the devolved nations and a wider range of expert opinion from the public, private and charitable sectors. I am particularly grateful for the inspired leadership of Paul Boyle in progressing this work, the huge contribution of Peter Elias in researching and preparing the final report and the tireless support of Vanessa Cuthill who kept us all on our toes.

I commend the recommendations of the Taskforce to Ministers in the Department for Business, Innovation and Skills and the Cabinet Office and I hope that our work will enable the government to grasp the opportunity that currently exists to make rapid progress in this area.

Sir Alan Langlands
Chair, Administrative Data Taskforce

Preface

Routinely collected administrative data are a rich and largely untapped source of information for research and policy evaluation in the UK, the value of which continues to appreciate over time. Such data are often high quality, nationally comprehensive, provide information for long periods of people's lives, and are relatively inexpensive to exploit, compared to the costs of establishing specially commissioned surveys.

Some other countries in Europe have forged ahead to capitalise on the opportunities these data provide and there is a risk that the UK will be left behind, which would be disappointing given our reputation as world leaders in the establishment and analysis of large-scale secondary datasets. This report recommends the establishment of the UK Administrative Research Network which will provide a robust data management and governance arrangement for analysing de-identified routinely collected administrative data.

Despite the poor progress in the UK, what we propose here is not new – there are examples in the UK of administrative data being linked between government departments and used productively in research projects. However, the number of examples is too few, the time taken to get agreement to use such data is too long, inconsistent decisions are being taken within government departments concerning rules of access and, most frustratingly, the legislative framework provided to allow for linkages to be made across departments is cumbersome and inefficient. Our recommendations, including vital legislative changes, would allow research that is already technically feasible to be undertaken in a much more consistent, reliable and efficient manner. This would be of huge value to academic research, but would also benefit research and policy evaluation within government departments, whose researchers are also constrained by the existing arrangements.

What we propose is in line with the 'open data' ambitions of the current government, as expressed in the recent White Paper, and could make a major contribution to the growth agenda. However, while we recognise the significant opportunities for academic and government research we also recommend proactive engagement with public on these issues. Such engagement to date suggests that once the public are made familiar with the strict controls that are put in place to analyse data from which the personal, identifiable information has been removed they support the use of these data for projects that are in the interests of the public good.

The Taskforce was initiated by a group of research funders and our final report provides an independent set of recommendations that we present to government. We are particularly appreciative of the input from those government officials who sat on the Taskforce in their capacity as data experts, as their insights helped us shape an ambitious system which they felt could be practically delivered. Their views should not be taken to represent the opinions of those departments, nor should their active contribution to the production of this report be taken to imply acceptance by government of the recommendations. It is now for government to formally consider the recommendations set out here and to respond appropriately. Finally, the funders would like to express our indebtedness to Sir Alan Langlands who chaired the Taskforce.

Professor Paul Boyle
Chief Executive, Economic and Social Research Council

Executive summary and recommendations

National administrative data such as social security, tax and education records relating to individuals or organisations have the potential to provide a robust UK-wide evidence base for research and policy evaluation. The Administrative Data Taskforce finds that access to and use of such data for research purposes in the UK has been difficult, due mainly to the concerns that data holders have about the possibility that information that identifies individuals could enter the public domain or because of legal restrictions they face on the uses to which such data can be put.

The Taskforce recognises that the UK has the opportunity to be a world leader in research using de-identified administrative data routinely collected by government departments, agencies and statutory bodies. Such data, made accessible for research in ways that prevent the identification of individuals, will inform a wide research agenda relating to the health, wealth and security of the UK population, guiding the development, implementation and evaluation of related policy interventions. The efficiency gains, in terms of both the high costs of alternative research resources and the speed with which findings can be generated, are likely to be considerable.

Realising this opportunity requires improvements in procedures for access to and linking between such data. This entails not just the development of a safe, secure and efficient system for linking, managing and analysing administrative data, founded on secure technologies, but on further building of trust between data subjects, data providers, researchers and all other relevant parties. The new system must adopt the highest international standards of governance, professional practice and public engagement. New legislation is needed to enable the efficient sharing and linking of data; government and the relevant funding agencies need to resource the new system to ensure its integrity, sustainability and utility.

The following recommendations of the Administrative Data Taskforce provide the framework to support these goals. In their entirety, they will create the **UK Administrative Data Research Network** – a collaboration between government departments and agencies, research funders and the research community that will reach across the UK to facilitate research based upon linked administrative data.

R1 An Administrative Data Research Centre (ADRC) should be established in each of the four countries in the UK

R1.1 The ADRCs will be responsible for commissioning and undertaking linkage of data from different government departments and making the linked data available for analysis, thereby creating new resources for a growing research agenda. Analyses of within sector data (e.g. linking medical records between primary and secondary care) and linking of data between departments for operational purposes may continue to be conducted by the relevant government departments and agencies.

R1.2 Each ADRC must be a state-of-the-art facility for research access to de-identified administrative data. While the detailed organisation and structure of each centre may reflect national variations in access arrangements, the fundamental common feature of each ADRC will consist of a **secure access facility** that meets the most stringent international standards. These should be attractive research environments (i.e. powerful hardware and analytical software should be available together with access to relevant metadata supplied by the data owners). Access to data and methodological and statistical support should be free to *bona fide* researchers who have gained accreditation status (see recommendations 3.1 to 3.3).

R1.3 Personal identifiers (names, addresses, precise date of birth, national insurance numbers, etc.) attached to administrative data records will not be available to, or held in, the ADRCs; hence, both ADRC staff and researchers accessing data through ADRCs will not have sight of such personal identifying information. Linkage will be achieved through the use of third parties who have the expertise to provide secure data linkage services for matching personal records from existing data systems.

R1.4 The ADRCs must maintain a full **audit trail** of all activity relating to data access and linking. This should include the establishment of monitoring and performance indicators. They must operate output control systems that provide for the vetting of all research outputs for confidentiality issues.

R1.5 The ADRCs must have the capacity to **conduct original research** using these data, as well as exploring issues relating to data linkage methods, the quality of

linked data, its coverage of specific populations and its suitability for particular research purposes. In addition to their own research function, the ADRCs will engage in training and capacity-building, supporting researchers who gain permission to undertake research on linked administrative data. By these means the ADRCs will contribute to the development of good practice for research access to de-identified personal data.

R1.6 Access to data held in the ADRCs by accredited researchers will be possible using three approaches. For all of these, **no individual-level records will be released from the ADRCs**. First, researchers can visit the ADRC secure data access facility, where their analyses of the relevant data sub-set will be overseen by the ADRC support team. Second, researchers can submit statistical syntax to the ADRC support team who will run the analysis on the dataset on behalf of the researcher (results would be thoroughly checked before return). Third, remote secure data access facilities may be established which allow virtual access to datasets held in the ADRCs. With the latter approach, no data would be transferred to these remote safe settings, which would use state-of-the-art technologies[1] and apply rigorous international standards, equivalent to those used in the ADRCs themselves, to provide a secure environment for researchers to undertake their analyses.

R1.7 A UK Governing Board will be established to provide the governance structure for the ADRCs. Together with the directors of the ADRCs, representatives from the funders and international experts in the field of administrative data use, membership of the Governing Board should include senior representatives from the relevant government departments, agencies and devolved administrative bodies that provide access to their data through one or more ADRCs. At least one lay member will also be appointed.

R1.8 The Governing Board will perform two key functions. At the strategic level it will have a focus on *leadership and enablement* in order to promote and facilitate safe research on administrative data for public and policy benefit. It will commission work and co-opt members if necessary to provide expertise in areas relating to ethical standards, international experience, linkage methodology (including linking administrative data to existing surveys and longitudinal studies), safe setting security, legal and other relevant issues.

In terms of its day-to-day work, a sub-group of the Governing Board will liaise closely with government departments and agencies to approve requests for access to and linkage between administrative datasets, as well as linkage to other non-administrative datasets. The Governing Board will also work to encourage collaboration across the ADRCs, and with other research data centres. To assist with this collaborative approach and to encourage cross-national working, the ADRCs will report on their activities on an annual basis to the Governing Board.

R1.9 The Governing Board will report on an annual basis to a body responsible to the UK Parliament which will monitor progress. We recommend that the UK Statistics Authority or another similar body fulfils this role. The report will include a set of relevant performance indicators and will detail achievements; obstacles encountered and proposed developments for the following year.

R1.10 An important part of the work of the ADRCs will be to ensure that proposals for access to and linkage between administrative datasets are legal, viable, and technically feasible and that the research skills necessary to conduct the research efficiently are available. Information about the ADRCs, including access arrangements, will be managed through an **Information Gateway**. This single web portal will be used to manage applications to use data and will provide detailed information about the administrative data that are available, the metadata attached to each of these datasets, and the results and impact of studies conducted through the ADRCs. The Information Gateway will likely be managed by one of the ADRCs.

R1.11 The ADRCs will provide access, free of charge, to government administrative data by publically-funded researchers, including those funded by or working on behalf of charities and the third/voluntary sector.
The ADRCs will not handle commercial data, or consider private sector requests for access to and linkage between administrative datasets held by public sector bodies. However, the Taskforce recognises that there could well be potential benefits that derive from private sector data and related research interests. The Governing Board will, at an early stage, investigate guidelines for access and linkage by private sector interests, as well as commissioning public engagement work on this topic.

[1] As provided at remote access points for the ONS Virtual Microdata Laboratory and the ESRC Secure Data Service.

It will evaluate the public benefits that will derive from specific requests for access, the need to ensure transparency in the research process, balancing these against the potential risk of reputational damage to data holders that might arise from the public perception of specific requests for access. In so doing, the Governing Board must note that there will be no derogation of the authority that government departments have in deciding how their data might be used in specific projects.

R1.12 An external review should be conducted after these arrangements have become embedded. This review would examine conditions of access for different groups of research interests, both public and private, with the aim of ensuring that these have evolved in a fair and open manner and with due regard for the views of the public.

R2 Legislation should be enacted to facilitate research access to administrative data and to allow data linkage between departments to take place more efficiently

R2.1 Government departments, Local Authorities, agencies and other public bodies face different legal restrictions on the nature of the access they might provide for research using data they control. In such cases, specific legal gateways have been established to resolve this issue. However, recent experience demonstrates that current link-specific gateway legislation is both cumbersome and inefficient. The Taskforce recommends that primary legislation should be sought to provide a **generic legal gateway** for research and statistical purposes that enables efficient access to, and linkage between, administrative data held by different government departments, agencies and other statutory bodies.

R2.2 An agreed set of ethical standards should be produced, drawing on well-established ethical guidelines and covering the research uses to which administrative data (and administrative data linked to other types of data, including surveys) may and may not be put.

R2.3 In situations where linkage is proposed between large and de-identified datasets, consent for linkage is not required under the Data Protection Act 1998. Where linkage involves the addition of administrative data to information collected by survey methods, it is both practicable and desirable to request consent for linkage from data subjects, even though the linked data will be de-identified prior to analysis. A common approach to the method of obtaining consent will be developed which will improve the efficiency of consent procedures and permit wider sharing of such linked data for research purposes.

R3 A single UK-wide researcher accreditation process, built on best national and international practice, should be established

R3.1 All researchers wishing to analyse administrative data through the ADRCs should be required to gain accreditation status.

R3.2 A single accreditation process needs to be designed which builds on best national and international practice and is acceptable to all UK holders of administrative data.

R3.3 This accreditation process will require short course training including: best practice methods for working in a secure data environment; legal and ethical issues related to the use of individual records; methods of disclosure control; and an introduction to the variety of types of routinely collected administrative data.

R3.4 To retain accreditation status, researchers should be required to undertake an online update course on a regular basis, possibly annually.

R3.5 Where there is no recourse to legal penalties for any breach of data access conditions, accreditation will be accompanied by a set of sanctions to be applied in a proportionate manner to researchers and/or their institutions in the event of any such breach of the conditions of access to administrative data.

R4 A strategy for engaging with the public should be instituted

R4.1 The Taskforce recognises the need to ensure that members of the public are aware of these developments in access to administrative data for research purposes, and that public concerns are given due attention. A strategy should be developed which will provide readily accessible information about the benefits of improved access to and linking between administrative data, and the measures being enacted to minimise risks of disclosure and to prevent inappropriate use of such data. The strategy should encompass procedures for raising public awareness about

the need for research based on administrative data, the nature of public consultation and the roles that members of the public could play in decision-making regarding the administrative data to be accessed and linked.

R4.2 To help develop this strategy, the Governing Board, which will include lay members, should steer the plans for public engagement prepared by each of the Administrative Data Research Centres (see recommendation 4.4), ensuring that there is no unnecessary duplication of efforts and suggesting ways in which these activities associated with these plans may be undertaken in the most efficient and effective manner.

R4.3 To assist with its implementation the Taskforce recommends that each of the proposed ADRCs should appoint **a public engagement and communications officer**, to lead engagement with a wide audience and promoting dialogue about the research benefits that could accrue, the safeguards that are required to prevent any misuse of data and how both could be effectively and collaboratively achieved.

R4.4 The ADRCs should produce plans for public engagement. They should collaborate to plan and hold public events to explain the work they are doing and to generate debate about the academic and broader social and economic benefits that derive from research using administrative data and the measures taken to ensure that the identities of individuals cannot be revealed.

R5 Sufficient funds should be put in place to support improved research access to and linkage between administrative data

R5.1 Funds should be secured to:

- establish the ADRCs, including technical and workforce requirements;
- support the data linkage activities of trusted third parties (organisations holding personal data which assist with secure data linkage procedures);
- support the development of the UK-wide researcher accreditation process and the provision of associated training courses;
- support the operational and strategic work of the Governing Board, including any legal work that is required, and the costs of the planned external review;

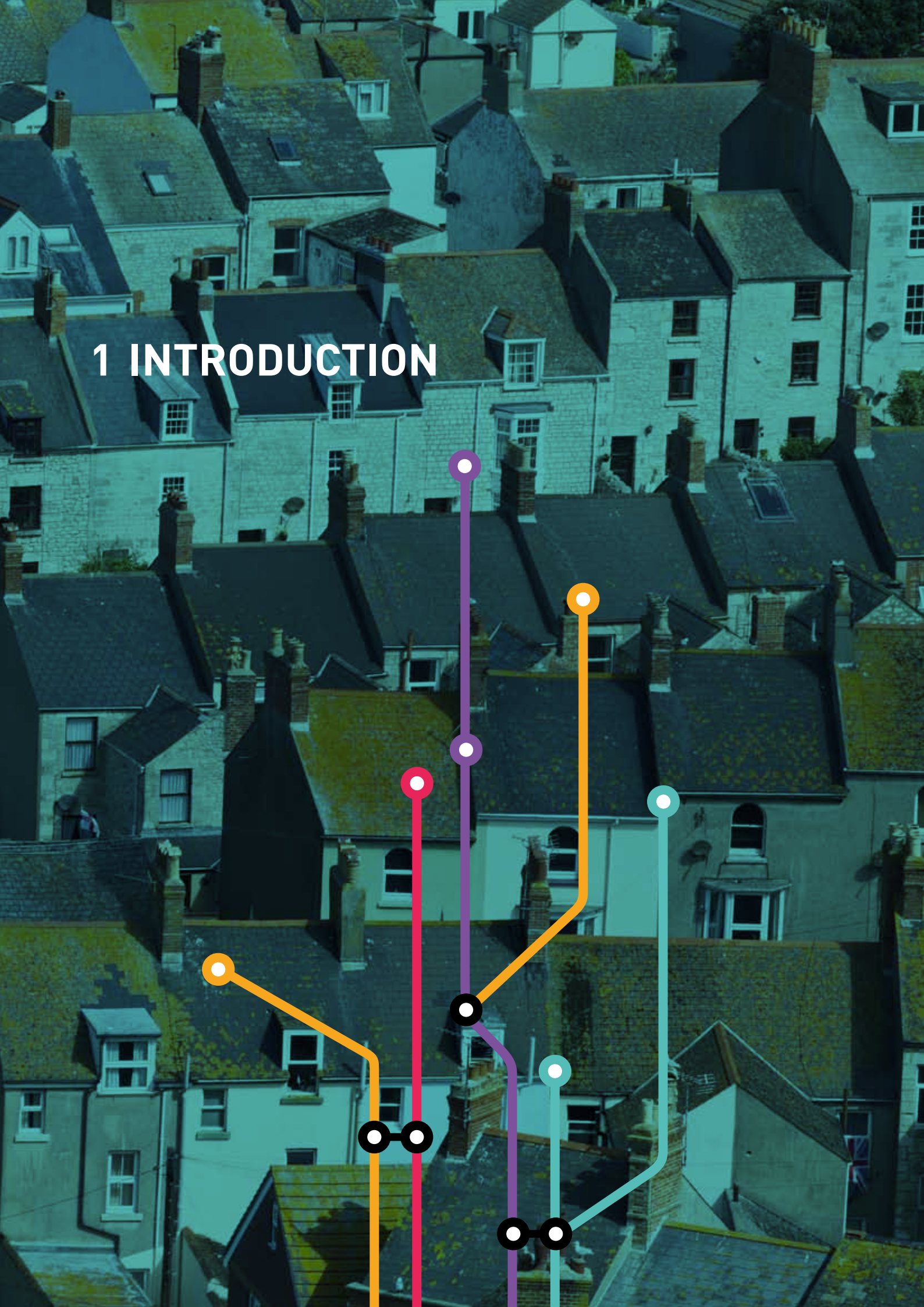
- support the independent auditing of the security procedures of the ADRCs;
- support for key activities (such as data retrieval, the creation of appropriate metadata, and the agreement on and implementation of agreed standards in data management) at the interface between the relevant government departments and the UK Administrative Data Research Network;
- assist Higher Education Institutions with the installation of secure rooms and the necessary equipment and staffing resource to provide virtual remote access to ADRCs.

R5.2 Funds to provide for the recommendations proposed in this report (comprising of the ADRCs, the governance structure, legal developments, accreditation and training, original research conducted within the ADRCs and the interface with government data holders) should be sought from an appropriate mix of interested parties including research councils, higher education funding councils, charities, and government.

The UK Administrative Data Research network

The various elements within this plan – the funders, the ADRCs, the Governing Board and links with the holders of administrative data – will combine to form a new data resource that will help position the UK at the forefront of research based on linked administrative data. Together they form the **UK Administrative Data Research Network**.

1 INTRODUCTION



1.1 The research value of administrative data

National administrative data collections held by government departments or agencies that relate to persons and/or organisations have the potential to provide a robust UK-wide evidence base that would contribute a rich new resource for research and policy making and evaluation. Improving access to and linkage between administrative datasets for research and statistical purposes would have demonstrable effects on economic growth and would help us respond more effectively to challenges related to the health and well-being of people. Making better use of these under-utilised resources will provide efficiency gains through the re-use of existing data, reduced reliance on more expensive methods of data collection and will speed the production of policy-relevant research. It sits neatly with the government's transparency agenda and is identified as a key commitment in the recent *Open Data* White Paper (Cabinet Office 2012).

The value that could be derived from such a resource relates to the policy relevant research it enables, examples of which include:

- addressing social mobility – by linking data on education, training, employment, unemployment, incomes and benefits;
- researching causal pathways over the life course – linking data on education, health, employment, incomes and wealth;
- comparative analysis of access to and the provision of social care support for the elderly;
- informing policies designed to tackle poverty – linking data on housing conditions, health incomes and benefits;
- constructing indicators of parental employment, social background, childcare and relating these to the provision of social care for children;
- linking data on (re)offending behaviour, incomes, benefits and health – exploring the role of poor mental health.

In addition to linking administrative data together across government departments, value can also be gained from linking administrative data to other studies, including ongoing longitudinal and other surveys. Linkages of this type have considerable potential for reducing the burden on respondents to such surveys and for improving the quality and extent of the information they provide.

> "IMPROVING ACCESS TO AND LINKAGE BETWEEN ADMINISTRATIVE DATASETS FOR RESEARCH AND STATISTICAL PURPOSES WOULD HAVE DEMONSTRABLE EFFECTS ON ECONOMIC GROWTH"

The United Kingdom is lagging behind some other countries in these respects. The Scandinavian countries and the Netherlands have made use of their extensive population registration statistics and other administrative data to facilitate important research in an efficient manner[2]. The United States Census Bureau has established a network of Research Data Centres which provide secure access to restricted census and survey microdata[3]. The UK must take steps to unlock the potential of its rich array of administrative data to become a world leader in terms of the research use of such data.

1.2 The Administrative Data Taskforce

Despite their considerable value as research resources, access to and linking between relevant administrative datasets has often been inhibited by issues relating to the legality of reuse and linkage for research and policy purposes. In some instances, research plans have been abandoned after funding for research has been agreed, or researchers have had to make use of data from other countries when faced with formidable obstacles to gain access to administrative data for the UK. The end result is that research opportunities that have clear public benefits to the UK are being missed. These are not new problems, with government departments urged to take action to address these issues some years ago[4]. Box 1 provides examples of some of the access problems from among many recorded by researchers and their funders[5].

[2] For further discussion and links on this topic see http://www.adls.ac.uk/adls-resources/guidance/introduction/

[3] See http://www.census.gov/research/data/restricted_use_microdata.php

[4] Thomas and Walport (2008) *Data sharing review* and the associated *Review of data handling procedures in government* (http://www.cabinetoffice.gov.uk/sites/default/files/resources/final-report.pdf)

[5] The Taskforce asked the Administrative Data Liaison Service to gather evidence from researchers and funders of cases where they had experienced problems with applications to access or link administrative data. These examples were typical of over 30 responses received.

Box 1: Examples of access and linkage problems experienced by researchers and research funders

Example 1: A researcher was requested by the Chief Medical Officer (CMO) for Wales to carry out research into the factors underlying excess winter mortality using the Secure Anonymised Information Linkage (SAIL) data bank, which has a facility to pseudonymise individual and household level data using an NHS trusted third party pseudonymisation service. This effectively ensures that all data viewed by researchers are anonymised. The study was to be a total population cohort with health data on individuals nested within households. Housing characteristics data, held by the Valuation Office Agency, were necessary for the study and the CMO for Wales requested that such data be pseudonymised and placed in the SAIL databank to facilitate this study. The request was made in April 2010 and after considerable efforts by HM Revenue and Customs (HMRC) to try and find a way to allow them to provide the information in a way which complied with their statutory constraints on data sharing it was eventually refused by HMRC in February 2012.

Example 2: We (a research funding organisation) have cases from both the Department for Education (DfE) and the Department for Work and Pensions (DWP) where access was agreed in principle along with clear background statements about the types of data to be linked (small area data, un-grouped benefit or income data). In one of these cases permissions were already given by survey respondents and it involved links between survey data and administrative data (drawn from the National Pupil Database). The actual steps needed to make access happen took over ten months. In the other case, involving links to data held by DWP, this took more than two years. This not only ties up research time but also must be inefficient. The planned research addressed issues where the departments were genuinely interested in the findings for various policy discussions but the delays meant the projects ran the risk of not being completed in time to inform government's own discussions, and with much less wider public debate, of the policy issues in question.

These examples do not clarify why the problems arose, which may have been made clear to the researchers concerned. What they do indicate is that there are problems with access to and linkage between government administrative datasets which impede valuable research. To realise the benefits that arise from improved access to administrative data, including better and more up to date information to inform policy developments, research funders, working with government departments formed the Administrative Data Taskforce to identify the factors inhibiting more widespread use of administrative data for research in the public interest and to make recommendations for improvements.

The terms of reference for the Taskforce are given at Appendix 1. The remit of the Taskforce relates specifically to improving the use for research of government administrative data. It does not extend to private sector data, or the provision of access to government data by the private sector. This is an important issue which will have to be addressed in due course and the Taskforce has borne this in mind in making its recommendations, particularly those which relate to the creation of new data infrastructure. Further discussion of this issue is given in section 2.7.

The Taskforce recognises the difficulties that many data holders have faced in responding to issues related to requests for access to their data for research and statistical purposes. Identifying legal pathways, ensuring confidentiality, designing robust methods for linkage to other datasets and undertaking such linkages are complex processes. Their effective management often places considerable strain on the limited resources that may be available to respond to such requests. This in turn has led to inconsistencies in decision making, variations in the interpretation of legal constraints, undue delays in access and different degrees of willingness to share data with academic and other researchers, all of which have contributed to the potential loss of important research opportunities.

Despite all these difficulties, some important examples of the public benefits deriving from cross-departmental data linkage are now becoming evident and demonstrate how, with sufficient leadership, motivation and resources, the hurdles to the development of powerful new research resources can be overcome. Box 2 illustrates this with

examples deriving from linkage across government departments. The issues that the Taskforce faces relate to the wider application of this experience across the full range of administrative datasets and the creation of the conditions for effective governance and resourcing of these types of activities.

The recommendations of the Taskforce address these issues. They are organised and presented in the following five sections of this report. These relate to: the establishment of new Administrative Data Research Centres (ADRCs) and associated governance arrangements; researcher accreditation and training in the research use of specific data resources; legal and ethical issues concerning access and linkage between different administrative datasets; the need for transparency and public engagement; and the resourcing of these recommendations.

Box 2: Sharing data between public authorities and with research bodies – two examples

Employment and benefit outcomes for ex-offenders
The Ministry of Justice (MoJ), the Department for Work and Pensions (DWP) and HM Revenue and Customs (HMRC) shared data on offenders, benefit claimants and employees to analyse employment and benefit outcomes for offenders, with the aim of supporting policy development.

Full legal and ethical approval for the data-sharing project was obtained in December 2010. The agreed administrative data was shared and successfully matched in early 2011, resulting in a dataset of approximately 3.6 million unique offenders with 40 million rows of sentencing, employment or benefit spells over the period 2000 to 2010.

Initial findings from the linked data were published in November 2011, to support policy development on the links between offending, employment and benefits and to demonstrate the potential of the improved evidence base (http://www.justice.gov.uk/downloads/statistics/mojstats/offending-employment-benefits-emerging-findings-1111.pdf).

The data-share has been used to develop a number of specific policy developments including the introduction of mandatory Work Programme provision for some ex-prisoners, a pilot looking to pay Work Programme providers for improvements to reoffending outcomes, and is enabling the MoJ and DWP to consider much wider impacts and appraisals of policies than had hitherto been possible.

Tracking progress through education and into the labour market

The Social Mobility Transparency Board was announced by the Deputy Prime Minister on the 22nd May 2012. The aim of the Board is to match and share a much wider range of data on progression through education and into the labour market, to underpin research, policies and initiatives outside government to improve social mobility. Three specific issues that are being considered are;

- Increasing the amount of linked education data that can be shared
- Linking education data with HMRC earnings and employment information
- Sharing Student Loan Company Repayment Information

If researchers are able to track learners throughout the education system and into the labour market this could lead to much better information on social mobility, advice and guidance, learning outcomes and graduate earnings by course and institution.

Through sharing Student Loan Company Repayment Information there could be a greater understanding of the Student Loan System as well as pathways through learning and the resulting variation in propensity for individuals to repay.

2 ESTABLISHING ADMINISTRATIVE DATA RESEARCH CENTRES

2 Establishing Administrative Data Research Centres

Various national and international models of access to national administrative datasets have been explored by the Taskforce (see appendix 2). These range from 'in-house' secure data access facilities (e.g. the HMRC Datalab) to remote access secure data facilities such as those developed by the Economic and Social Research Council (ESRC) Secure Data Service and the Office for National Statistics (ONS) Virtual Microdata Laboratory. New centres are currently being established in the UK for improved access to specific types of administrative data[6]. While the location and oversight of these facilities varies – some based in national statistical offices, some in academic settings – common to all are strict governance arrangements and adherence to standards for secure and safe access to microdata records for research.

The Taskforce recognises that a variety of models of access will continue to operate in the short term. However, most are not well resourced to cope with the step change in demand for access to administrative data by the research community that has arisen. Additionally, linking different types of administrative data for individuals – a powerful way of developing research data resources – remains problematic for various reasons explored in this report. What is required is a UK system for linking data and providing research access to linked data, together with an overarching governance structure that helps data holders to make consistent decisions about data linkage and access for research.

To achieve consistency across the UK in protocols for data linking, and to provide data holders with the assurances that procedures for data linking and research access are safe, secure and legal, it is proposed that a number of new centres should be established. The structure and purpose of these Administrative Data Research Centres (ADRCs) is outlined in this section. Given the different pace at which access and linkage to administrative datasets is taking place in the countries of the UK, and the existence of different datasets in each country, it makes sense to take advantage of the situation by locating at least one physical Administrative Data Research Centre in each country of the UK. While each centre will have its own staff and premises, they would build upon, take advantage of, and extend existing structures and expertise;

- in Scotland, work is already underway to develop a centre as part of the Scotland-wide Data Linkage Framework and the Scottish government has already pledged funds to help establish such a system. An ADRC could build on best practice from the experience of the Scottish Health Informatics Programme (SHIP) and the Scottish Longitudinal Study (SLS);
- in Wales there are similar plans, and the establishment of the ADRC would likely build upon the experience of the Secure Anonymised Information Linkage (SAIL) Databank, extending beyond its existing primary focus on health data;
- in Northern Ireland the ADRC could build upon the work of the Northern Ireland Statistics and Research Agency to establish the Northern Ireland Longitudinal Study (NILS);
- In England progress is less well coordinated and the ADRC would be commissioned to take advantage of knowledge and experience in administrative data access and linkage acquired by various research groups.

In addition to the personnel needed to operate each ADRC, resources will be required at the interface with the relevant government departments to undertake data extraction, work on data standardisation and develop metadata (information about data). The resources required will vary depending on the workload incurred by each department including the number of requests for data, and the legal standing of the department which may influence the process of data access and linkage.

The Taskforce views four ADRCs, one established in each country of the UK, as a minimum workable objective. Additional ADRCs may be established for specific purposes and with earmarked funding. For example, data linkage associated with any plans for a new approach to data collection in the decadal Census of Population may require a dedicated facility. In such instances it would be efficient for them to take advantage of the proposed governance, accreditation and training arrangements and procedures described in this report.

[6] Examples include the four health-focused e-Health Centres of Excellence (Dundee, Manchester, London and Swansea) established by a consortium of government and research funding agencies and the Ministry of Justice secure data access facility for research access to judicial and related data.

2.1 How would these Administrative Data Research Centres work?

In terms of the models of access shown in appendix 2, the Taskforce has identified an appropriate form of model 3 (trusted third party indexing) as providing the most trusted, robust and secure approach to research access to and linkage between administrative datasets. While the exact specification of each ADRC will depend to some extent on national circumstances, in the view of the Taskforce there are certain minimum requirements that must be met to achieve the appropriate security whilst providing a productive research environment. These are as follows:

- the ADRCs will be required to maintain state-of-the-art secure data access facilities and must operate linkage procedures that do not allow their staff or researchers to construct or have sight of personal identifying information on individual administrative data records. It is therefore anticipated that the approach underpinning the data linkage process will be similar to that used by the Scottish Longitudinal Study (SLS), where personal identifying information is not held in the ADRC, but is matched through a third party service, such as the National Health Service Central Register (see box 3 for an example of how this works);
- they must provide the level of data security consistent with the security classification of the data being analysed;
- the models of access they employ must be flexible enough to cope with varying access conditions relating to the legal powers that govern the sharing of data by government departments;
- they would undertake their own original research programme of work, as well as work on improving understanding of the research value and quality of administrative data, particularly where different datasets had been linked;
- they would be resourced to provide a data management and statistical analysis support function for external researchers accessing the data;
- they would provide an attractive research environment (i.e. powerful hardware and analytical software should be available together with access to relevant standardised metadata which should be supplied by the relevant government department);
- they must have record keeping systems which yield a full audit trail of all activity relating to data access and linking. Each will operate output control systems which provide for the vetting of all research outputs for confidentiality issues;
- they would be subject to regular external audit of their data security protocols and systems;
- they would establish and monitor performance indicators that evaluate the time it has taken for researchers to gain access to data, the gains in efficiency for departments and the increased research outputs from the new system.

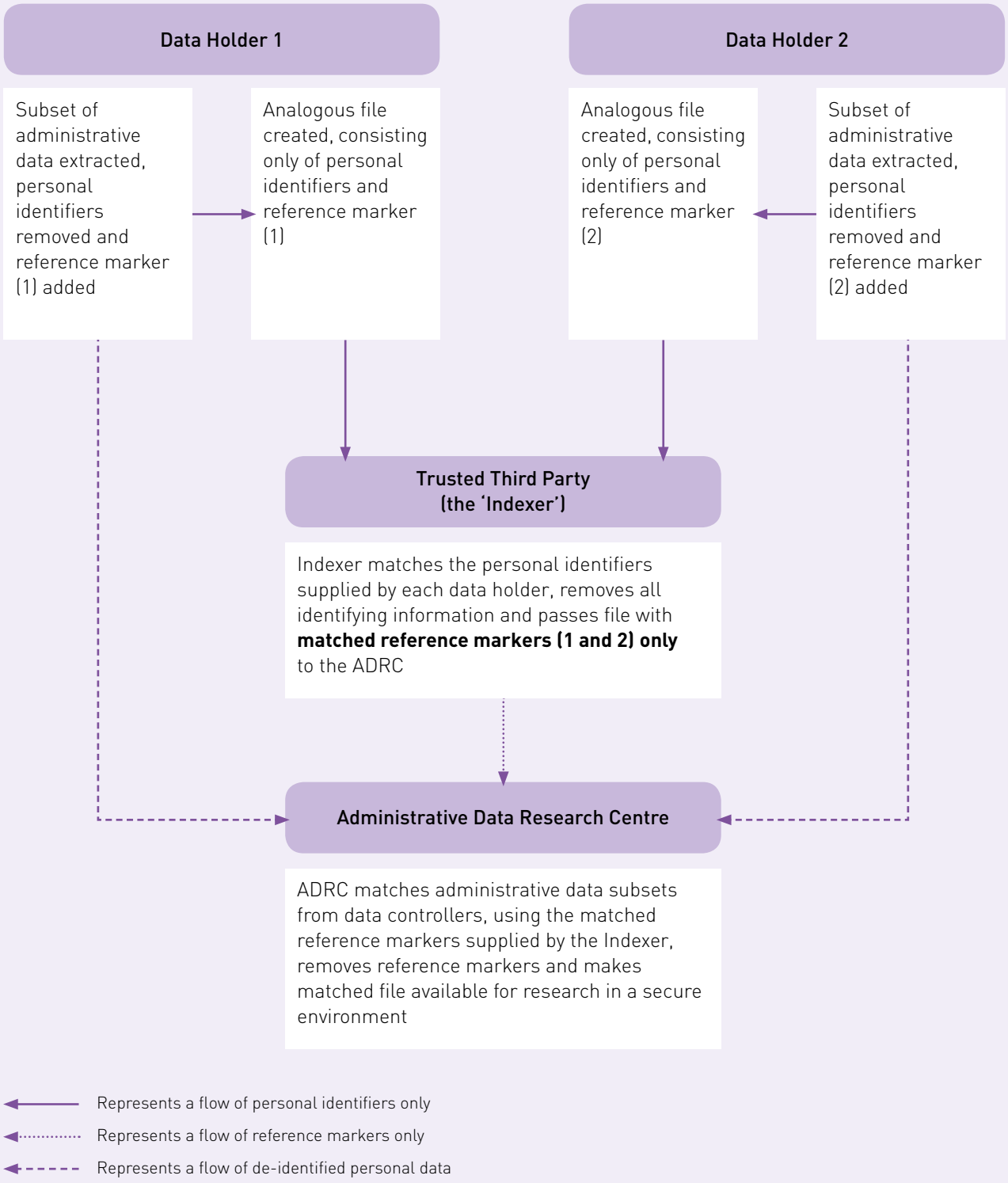
Trusted third party indexing has the advantage that the ADRCs and the researchers who make use of data held by the ADRCs never have access to or sight of personal identifying information. Additionally, the trusted third party never sees the information about identified individuals that forms the data holder's administrative data. Box 3 shows how the process works, illustrating a situation where data from two holders is matched and linked for research purposes.

2.2 Governance of the Administrative Data Research Centres

Essential features that all ADRCs must have in common are the governance arrangements that should be in place to promote and guide their activities.

The four national ADRCs will be independent of government, and a single UK **Governing Board** will be established to provide the governance structure. Together with the directors of the ADRCs, representatives from the funders and international experts in the field of administrative data use, membership of the Governing Board would include data holders from the relevant government departments, agencies and devolved administrative bodies that provide access to their data through one or more ADRCs. At least one lay member would also be appointed.

Box 3: Data Linkage – an example of de-identifying records through the use of a Trusted Third Party



2.3 The functions of the Governing Board

The Governing Board would have strategic and day-to-day functions. Strategically, it will focus on *leadership and enablement* in order to promote and facilitate safe research on administrative data for public and policy benefit. It should commission work and co-opt members if necessary to provide expertise in areas relating to ethical standards, international experience, linkage methodology (including linking administrative data to existing surveys and longitudinal studies), safe setting security, legal and other relevant issues. It will work to encourage collaboration and sharing of best practice across the ADRCs, and with other research data centres in the UK and internationally.

In terms of its day-to-day workload a sub-group of the Governing Board will work closely with data holders to progress and secure approval for requests for access to and linkage between administrative datasets, as well as linkage to survey and other datasets.

2.4 Reporting arrangements for the Governing Board

The strategic role of the Governing Board is to promote enhanced access to and linkage between routinely-collected administrative data held by various government departments and agencies. Its membership gives it cross-departmental, UK-wide representation. Given its proposed interdepartmental membership, reporting by the Governing Board to any one particular government department may not be appropriate. This, together with its UK remit, suggests that it should report on an annual basis to a body responsible to the UK Parliament which will monitor progress. We recommend that the UK Statistics Authority or another similar body fulfils this role.

The role of the body to which the Governing Board reports will include: approval of the Terms of Reference for the Governing Board; oversight of the appointment of Governing Board members; and receipt of an annual report from the Board. The annual report will include a set of relevant performance indicators as well as achievements, obstacles encountered and proposed developments for the following year.

2.5 Providing an Information Gateway to Administrative Data

Good research design, incorporating appropriate administrative data, requires that the research community should have knowledge about such data – its provenance, coverage, quality, and the conditions surrounding access for research purposes. These services would be delivered via an **Information Gateway.**

The Information Gateway would consist of a single website portal to the work of the four proposed ADRCs, responding to requests for information and advice about the use of administrative data for research. Applications to link data for research would be made through this portal. It would also build upon and extend the work of the existing ESRC-funded Administrative Data Liaison Service in the following manner:

- it would liaise with the four national ADRCs, as well as other centres which provide access to specific datasets or data types, providing examples of research based on successful data access and record linkage;
- it would facilitate networking and the sharing of knowledge and experience with the research use of administrative datasets between researchers and data holders.

In addition to these web-based services, the Information Gateway would also operate a researcher accreditation service in association with data holders. Such accreditation would include training to researchers and departmental data holders in the use of safe settings for data access, data disclosure and secure data management, as well as on the types of data that are available and basic and advanced methods for analysing them (Section 3).

The Gateway would be the first port of call for prospective researchers, providing them with the information, knowledge and credentials that they would need to gain access to and use a wide variety of types of administrative data for their research. By providing such a 'front-line' service to researchers, the Information Gateway would reduce the demands for information and access regularly made to data holders and to current research centres involved in data linkage.

The Information Gateway web portal service could be managed by one of the proposed ADRCs, but it would liaise with the three other ADRCs and relevant government departments and agencies.

2.6 Charitable bodies, the third/ voluntary sector and research using administrative data

Bodies with charitable status and the third/voluntary sector may wish to interact with the UK Administrative Data Research Network in a variety of ways. As funders, they could choose to support elements of the network through the provision of financial resources. They may also promote and support research within the academic sector which takes advantage of the access and linkage facilitates that the network provides. Additionally, charities and third/ voluntary sector organisations may themselves undertake research in the public good which is consistent with their aims and which makes use of network resources.

Charitable bodies and third/voluntary sector organisations will be afforded the same status as publically-financed research institutions. Where their involvement includes the provision of resources to support the network, consideration will be given to their representation on the Governing Board.

2.7 Private sector interests and data linkage

During the period of its formation, the UK Administrative Data Research Network will not consider requests for access to administrative data made by private sector organisations and will not permit linkage between public data and commercial datasets.

The recommendations made by the Taskforce relate only to access to and linkage between government administrative datasets for publically-funded research. These recommendations, when implemented, will create new and/or enhanced data for research and will help remove or reduce barriers that have impeded progress on, or prevented research from, being conducted. It is anticipated that this will stimulate interest in research using such data, from both the public and the private sectors, raising questions about research access by the private sector.

Commercial databases include large customer databases covering store purchases, financial transactions, mobile communications, social networking and internet-enabled search activities. Their linkage to administrative data held by public bodies could provide information to improve the efficiency of the operations of private sector organisations. Equally, access by publically-funded researchers to the types of data held by private sector organisations and their linkage to administrative data held by public sector bodies has the potential to inform research with strong public benefits. There are clear examples of the ways in which data sharing and linkage between public and private sector bodies can benefit both parties as illustrated in Box 4.

The Taskforce has considered this issue and identifies the need to consider the balance between access to government-controlled administrative data by private sector organisations for restricted research and access which yields strong public benefits. There is also a need to consider carefully the possibility that the public perception of requests for access and linkage to public sector administrative data by private sector organisations may lead to reputational damage, with negative repercussions for public sector agencies.

To achieve this balance, it is recommended that the Governing Board should, as a matter of some priority, review the range of interests in access to administrative data expressed by public and private bodies and establish guidelines for access according to the nature of the research. In so doing it will evaluate the public benefits that will derive from specific requests for access, the need to ensure transparency in the research process and the potential risk of reputational damage to data holders that might arise from the public perception of specific requests for access. It will engage with the public to ascertain the willingness for public administrative data to be shared with the private sector. This review would examine conditions of access for different groups of research interests, with the aim of ensuring that these evolve in a fair and open manner and with due regard for the views of the public. Note, again, that no government dataset would be used in any research project without the explicit agreement of the respective government department.

Box 4: Examples of the scope for and benefits of public / private data linkage

Energy Companies – identifying multi-occupied dwellings for the 2011 Census

One area which has already proven to be of value has been the Office for National Statistics' (ONS) use of a file built by electricity companies and containing the addresses of properties with multiple meters, to improve its address file for the 2011 Census. These quality improvements in areas difficult to address accurately provided one of many enablers for ONS to increase response rates considerably when compared to the last Census in 2001.

Energy Companies – seeking to target fuel poverty

A household is defined as being in fuel poverty if it needs to spend more than ten per cent of its income on fuel to maintain a satisfactory heating regime (usually 21 degrees for the main living area, and 18 degrees for other occupied rooms). Whether a household is in fuel poverty or not is determined by the interaction of a number of factors, but three stand out.

- The energy efficiency of the property
- The cost of energy
- Household income

Energy companies are required to meet government fuel poverty targets, and are interested in external datasets, especially from government itself. There is a general preference for individual-level data, and information on 'super-priority groups', such as low-income people on benefits. Fuel poverty is most prevalent amongst families with children; also sickness causes loss of income plus the need for more fuel at home. Improved access to administrative datasets, would help energy companies to meet these targets, and could contribute to the development of a better service for consumers.

2.8 Examples of administrative data research centres

What we propose in this report is not entirely new, as there are examples in the UK of administrative data being linked between government departments and used productively in research projects. Boxes 5 and 6 give examples of two important initiatives which have enabled data linkage and access for research purposes to be established in ways which maintain the confidentiality of personal data yet permit research to be undertaken at the microdata level.

Box 5

The Scottish Longitudinal Study (SLS)

The SLS is a large-scale linkage study which has been created by using data available from current Scottish administrative data sources. Over five per cent of the Scottish population are included. The SLS is the most advanced example of linking data across government sectors in the UK and the data held include those from the Census, vital events (birth, marriages and deaths) records, cancer registrations, hospital admissions and educational attainment and the school census. The current database holds almost 800 variables derived from the above sources and is set to expand further as the records from the 2011 Census of Population are linked in. Personal data are never held by the SLS team and the Scottish National Health Service Central Register is used as the third party indexing service. Linked records are available for analysis as de-identified individual-level data in either a safe setting, or via the submission of statistical syntax which is applied to the relevant data on behalf of the researcher. Various measures have been taken to ensure confidentiality whilst facilitating detailed research. Outputs are screened to ensure that no inadvertent disclosure of identities is possible.

Further details about these data and access procedures, as well as the research being conducted within the SLS, are available at www.lscs.ac.uk/sls/access.htm

Box 6

The Michigan Census Research Data Center (MCRDC)

The MCRDC is a joint project of the US Bureau of the Census and the University of Michigan. It enables qualified, confidentiality-abiding researchers with approved projects to access confidential, unpublished Census Bureau data under the provisions of Title 13, sec. 9 U.S. Code. The Census Bureau's Center for Economic Studies has developed and put into practice the concept of Research Data Centers (RDCs). The RDCs provide a secure Census Bureau environment where researchers may have limited access to confidential economic, demographic and public health microdata, with appropriate safeguards to protect data confidentiality. This controlled environment ensures that the Census Bureau's standards for ensuring the confidentiality of data by its census and survey respondents is rigorously maintained.

Source: http://www.isr.umich.edu/src/mcrdc/data.html

2.9 Recommendations

- An Administrative Data Research Centre (ADRC) should be established in each of the four countries in the UK. These ADRCs will be responsible for commissioning and undertaking linkage of data from different government departments and making the linked data available for analysis, thereby creating new resources for a growing research agenda. Analyses of within sector data (e.g. linking medical records between primary and secondary care) and linking of data between departments for operational purposes may continue to be conducted by the relevant government departments and agencies.

- Each ADRC must be a state-of-the-art facility for research access to de-identified administrative data. While the detailed organisation and structure of each centre may reflect national variations in access arrangements, the fundamental common feature of each ADRC will consist of a **secure access facility** that meets the most stringent international standards. These should be attractive research environments (i.e. powerful hardware and analytical software should be available together with access to relevant metadata supplied by the data owners). Access to data and methodological and statistical support should be free to *bona fide* researchers who have gained accreditation status (see recommendations 3.1 to 3.3).

- Personal identifiers (names, addresses, precise date of birth, national insurance numbers, etc.) attached to administrative data records will not be available to, or held in, the ADRCs; hence, both ADRC staff and researchers accessing data through ADRCs will not have sight of such personal identifying information. Linkage will be achieved through the use of third parties who have the expertise to provide secure data linkage services for matching personal records from existing data systems.

- The ADRCs must maintain a full **audit trail** of all activity relating to data access and linking. This should include the establishment of monitoring and performance indicators. They must operate output control systems that provide for the vetting of all research outputs for confidentiality issues.

- The ADRCs must have the capacity to **conduct original research** using these data, as well as exploring issues relating to data linkage methods, the quality of linked data, its coverage of specific populations and its suitability for particular research purposes. In addition to their own research function, the ADRCs will engage in training and capacity-building, supporting researchers who gain permission to undertake research on linked administrative data. By these means the ADRCs will contribute to the development of good practice for research access to de-identified personal data.

- Access to data held in the ADRCs by accredited researchers will be possible using three approaches. For all of these, **no individual-level records will be released from the ADRCs**. First, researchers can visit the ADRC secure data access facility, where their analyses of the relevant data sub-set will be overseen by the ADRC support team. Second, researchers can submit statistical syntax to the ADRC support team who will run the analysis on the dataset on behalf of the researcher (results would be thoroughly checked before return). Third, remote secure data access facilities may be established which allow virtual access to datasets held in the ADRCs. With the latter approach, no data would be

transferred to these remote safe settings, which would use state-of-the-art technologies[7] and apply rigorous international standards, equivalent to those used in the ADRCs themselves, to provide a secure environment for researchers to undertake their analyses.

- A UK **Governing Board** will be established to provide the governance structure for the ADRCs. Together with the directors of the ADRCs, representatives from the funders and international experts in the field of administrative data use, membership of the Governing Board should include senior representatives from the relevant government departments, agencies and devolved administrative bodies that provide access to their data through one or more ADRCs. At least one lay member will also be appointed.

- The Governing Board will perform two key functions. At the strategic level it will have a focus on *leadership and enablement* in order to promote and facilitate safe research on administrative data for public and policy benefit. It will commission work and co-opt members if necessary to provide expertise in areas relating to ethical standards, international experience, linkage methodology (including linking administrative data to existing surveys and longitudinal studies), safe setting security, legal and other relevant issues. In terms of its day-to-day work, a sub-group of the Governing Board will liaise closely with government departments and agencies to approve requests for access to and linkage between administrative datasets, as well as linkage to other non-administrative datasets. The Governing Board will also work to encourage collaboration across the ADRCs, and with other research data centres. To assist with this collaborative approach and to encourage cross-national working, the ADRCs will report on their activities on an annual basis to the Governing Board.

- The Governing Board will report on an annual basis to a body responsible to the UK Parliament which will monitor progress. We recommend that the UK Statistics Authority or another similar body fulfils this role. The report will include a set of relevant performance indicators and will detail achievements; obstacles encountered and proposed developments for the following year.

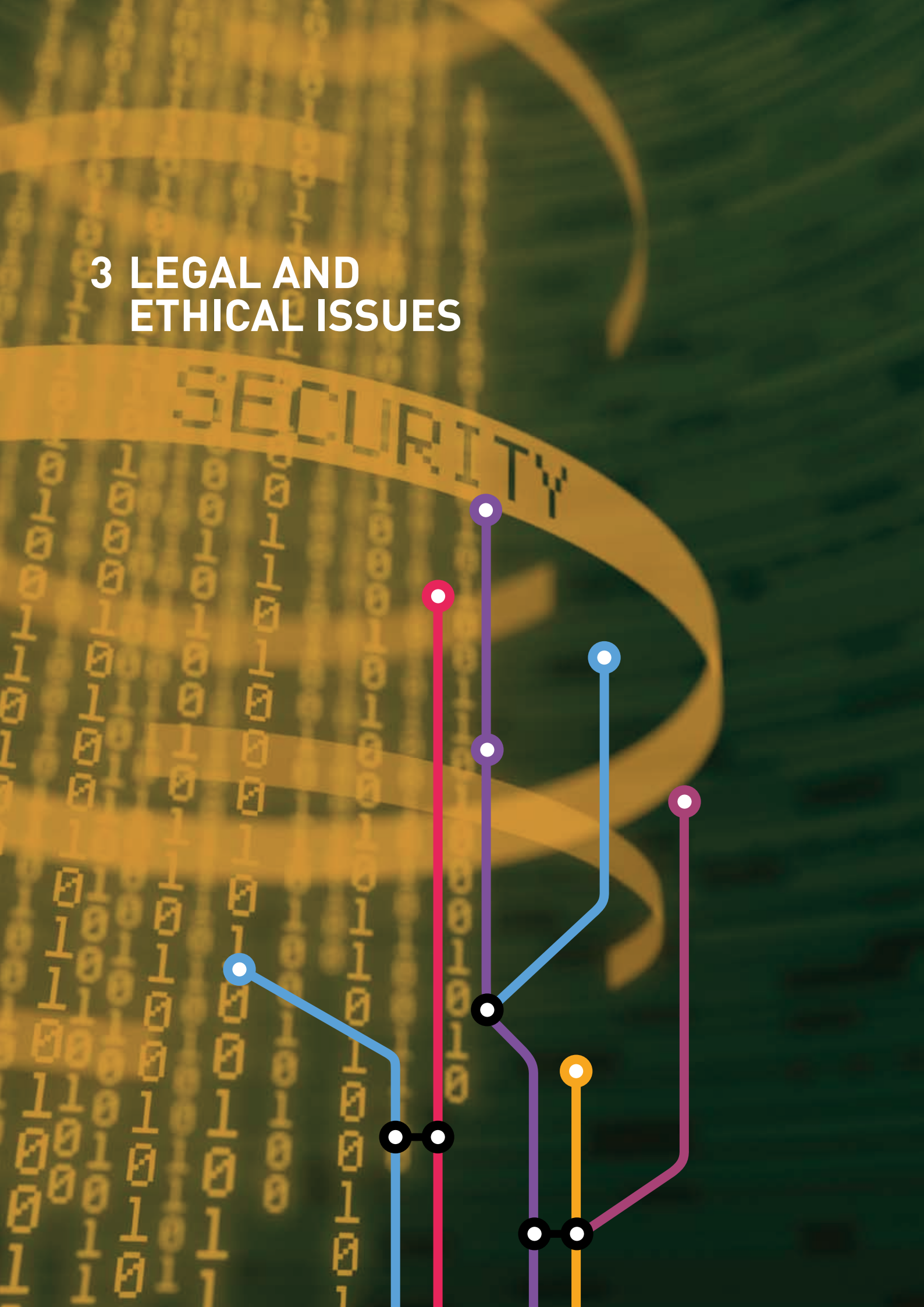
- An important part of the work of the ADRCs will be to ensure that proposals for access to and linkage between administrative datasets are legal, viable, and technically feasible and that the research skills necessary to conduct the research efficiently are available. Information about the ADRCs, including access arrangements, will be managed through an **Information Gateway**. This single web portal will be used to manage applications to use data and will provide detailed information about the administrative data that are available, the metadata attached to each of these datasets, and the results and impact of studies conducted through the ADRCs. The Information Gateway will likely be managed by one of the ADRCs.

- The ADRCs will provide access, free of charge, to government administrative data by publically-funded researchers, including those funded by or working on behalf of charities and the third/voluntary sector. The ADRCs will not handle commercial data, or consider private sector requests for access to and linkage between administrative datasets held by public sector bodies. However, the Taskforce recognises that there could well be potential benefits that derive from private sector data and related research interests. The Governing Board will, at an early stage, investigate guidelines for access and linkage by private sector interests, as well as commissioning public engagement work on this topic. It will evaluate the public benefits that will derive from specific requests for access, the need to ensure transparency in the research process, balancing these against the potential risk of reputational damage to data holders that might arise from the public perception of specific requests for access. In so doing, the Governing Board must note that there will be no derogation of the authority that government departments have in deciding how their data might be used in specific projects.

- An external review should be conducted after these arrangements have become embedded. This review would examine conditions of access for different groups of research interests, both public and private, with the aim of ensuring that these have evolved in a fair and open manner and with due regard for the views of the public.

[7] As provided at remote access points for the ONS Virtual Microdata Laboratory and the ESRC Secure Data Service.

3 LEGAL AND ETHICAL ISSUES



3 Legal and ethical issues

While the potential value of using administrative data for analytical purposes inside and outside government is well understood, the legal and ethical issues surrounding data access and linking are complex. There have been different approaches arising from the different legal provisions and policies applying to various categories of data. On occasion, data have simply not been made available even though there may have been a lawful solution.

3.1 The legal framework

When considering whether to share its data for any purpose a government department has to consider the powers available to it, along with any constraints which apply to those powers. The powers themselves may exist in such a way that they can only ever be used in limited circumstances. For example, powers created by legislation may only be available in particular situations specified in that legislation. By contrast, other powers may allow for greater flexibility, but that flexibility will not be unlimited.

Additionally, the wider framework within which government powers are exercised (e.g. the Data Protection Act 1998 and Human Rights Act 1998) may have an effect. Sometimes the effect will be a narrowing one, restricting further the powers that a government department has to share data for analytical purposes. However, this does not automatically follow and, as a general rule, the impact of the wider framework on whether an individual act of data-sharing is permissible is considered on a case-by-case basis.

Legislation in the area of data protection is likely to be revised as the European Union seeks to introduce a new Regulation. This could have repercussions for research access to data if specific derogations – or exceptions – from particular requirements for the use of personal data for historical, statistical and scientific research purposes are not clearly specified (see Box 7 opposite).

In assessing whether they are able to link data, departments will usually consider whether any of the following are available:

- **express statutory powers** – these are powers specifically given by, usually primary, legislation;
- **implied statutory powers** – these are powers not given on the face of legislation but which can be implied from legislation because without them a department would be inhibited in carrying out a legislative function in the way envisaged by the relevant legislation; or
- **common law powers** – these are powers which exist outside legislation. The most common expression of these powers for government is by reference to the Ram Doctrine[8], by which it is understood that a department is able to do the same things as a private citizen, subject to any constraints which may apply (for example in statutory provisions).

Not all of these powers are available to departments wishing to share and/or link data. For example, bodies such as HMRC and the Welsh Government derive all of their powers and duties from statute. HMRC does not have any common law powers, in relation to sharing data or otherwise. This means that, as well as having to observe express statutory restrictions placed upon its use of data (e.g. section 18 of the Commissioners for Revenue and Customs Act 2005), HMRC does not have a wider suite of powers outside of its legislative framework. Consequently, HMRC may agree to data linkage for studies that fall within its statutory remit, but proposals for linkage to HMRC data for studies which do not fall squarely in this remit will require a specific legal gateway or, where feasible, the informed consent of subjects of research.

Other departments, as emanations of the Crown, benefit from the Royal Prerogative, and are not subject to the same sorts of constraints. For these departments the common law, through the Ram Doctrine, provides greater flexibility. However this flexibility may be restricted where express statutory powers and gateways touch on the areas within which such departments operate. The existence of express statutory powers in a particular area can pose a question about whether the common law has any remaining role to play in that area or whether the 'field has been occupied' by the express power with the result that the common law power has effectively been ousted. For example, the Department for Education relies on its common law powers to disclose aggregated data to researchers in some areas. By contrast it has express powers to share individual pupil data through

[8] Expressed in a memorandum of advice given by the then First Parliamentary Counsel, Sir Greville Ram, on 2 November 1945, the doctrine contends that the Crown has all the powers of a natural person and, unlike a statutory body, does not need to point to any statutory power or authority for any action they may wish to take. In other words, a Secretary of State in a Crown government department may exercise any powers that the Crown has power to exercise, except in so far as he/she is precluded by statute from so doing, either expressly or by necessary implication.

the mechanism provided by the Education Act 1996 which sets the limits of data sharing of this sort. This complexity means that there are real questions to be explored about whether new proposals to share data for research purposes would be permissible within existing legal powers, whether express, implied or common law. Even where sharing is possible, the complexity can result in a lack of clarity and confidence in the legal position which, in turn can lead to cautious decisions.

Box 7

Proposed European Union Data Protection Regulation

In January 2012 the European Commission published a draft Data Protection Regulation with a view to replacing the existing Data Protection Directive and associated Member State legislation such as the UK Data Protection Act. Unlike a Directive, a Regulation takes direct effect rather than being transposed into Member State law. The Regulation is currently being considered and amended by the European Parliament and Council of Ministers before it is adopted.

Like the Data Protection Directive, the Regulation applies to the processing of personal data across a wide range of sectors, including research. The Regulation is designed to strengthen individuals' rights and increase harmonisation of data protection approaches across the EU. The Regulation introduces new requirements for the processing of personal data. These have raised concerns among many organisations and the UK government has said it will 'resist new bureaucratic and potentially costly burdens on organisations which do not appear to offer greater protection for individuals.'

The Regulation appears to provide a number of derogations – or exceptions – from particular requirements for the use of personal data for historical, statistical and scientific research purposes.

In order to qualify for these derogations, personal data must be processed in accordance with conditions set out in Article 83: personal data should not be used if anonymous data would be sufficient and, if possible, any identifying information should be kept separately from other information. This approach has been welcomed by many in the research sector since it provides a framework that balances the facilitation of research with the protection of the interests of research participants. It will be important to ensure that Article 83 and the associated derogations for research are protected as the Regulation moves through the legislative process. It is also important to note though that these derogations do not exempt research studies from all the requirements set out in the Regulation, some of which may pose difficulties for research unless amended.

It is vital there is clarity around how data protection legislation relates to de-identified or 'key-coded' data used in research. The Information Commissioner's Office has published an Anonymisation Code of Practice[1] that provides a helpful indication that de-identified data can be considered anonymous where certain conditions are fulfilled. However, the Regulation does not currently provide such clarity on the status of de-identified or key-coded data. Informal indications from the Commission suggest that their approach could be consistent with that of the Information Commissioner's Office, but it will be important that changes are made to ensure that the legislation is clear on this point and that de-identified data are regulated proportionately.

[1] Information Commissioner's Office (2012) Anonymisation: Managing Data Protection Risk.

3.2 Data linking and access – how the legal framework shapes the process

Where government departments involved in a potential data-linking project have different powers the legal solutions available are limited. That may in turn impact upon the arrangements for data linking and access.

Where a government department needs statutory powers to share data there is often a criminal sanction for unlawful sharing of data that relates to an identifiable person/legal entity, most commonly a prison sentence of up to two years. This has led to some degree of caution that may go beyond what is required by the legislation itself. This can apply to individual judgements around access to data, where possible solutions are not fully explored because of the perception of the barriers. These legal issues apply at both key stages of the data linkage and access procedures. At the point of linking information between two organisations each must decide whether it can lawfully provide the data. Each must consider whether it has a legal power to provide or receive, either under statute or common law. Where personal data are included, data protection and human rights issues, including purpose and proportionality, must also be considered. The existence of personal data does not necessarily rule out data linking but may do so. In some case security safeguards or privacy measures may be enough to render the linkage lawful.

The legal options at this stage may determine how data linkage and access is achieved. Consider the case where personal information is needed and one data holder does have a power to share such data (including in accordance with data protection and human rights rules), but the second data holder does not. The first data holder can give its data to the second, but not vice versa. In this case the former should share data with the latter and the latter (or a contracted data processor) should do the linking since they can more easily have lawful access to both datasets. Where the second data holder undertaking the linking was a department created by statute, it would still require a legal gateway to allow it to share the linked data with the first data holder or with another party, unless the research supported the second data holder's departmental functions.

In terms of granting access to the data, as a first step there should be consideration of whether the data can be made non-personal. If the information is non-personal, the constraints on departments when sharing data are fewer, but some departments are still unable to share non-personal data. If personal information is to be shared, wider considerations such as data protection and human rights will need to be taken into account. In such a case, consideration may need to be given to whether a contract is appropriate in order to ensure that sufficient safeguards are in place to address data protection or human rights requirements.

> "WHERE PERSONAL DATA ARE INCLUDED, DATA PROTECTION AND HUMAN RIGHTS ISSUES, INCLUDING PURPOSE AND PROPORTIONALITY, MUST ALSO BE CONSIDERED."

Note that the ability to use existing legal powers may vary between the two stages. Some departments known to have considerable restrictions on sharing data with external researchers have considerable legal powers to share data between each other (HMRC and DWP have such an express data sharing power). The linkage decision at the start of the project should consider both stages in order to determine the legal options that would have a considerable effect on the technical means of linking and giving access.

3.3 Legal gateways for data linking and access

Some government departments and agencies have established specific legal gateways which may permit data linkage but are more usually designed to facilitate data sharing between data holders. These gateways are enabled by statute and specify the nature of the data to be shared, by whom and in what circumstances. Examples include data from undertakings collected under the Statistics of Trade Act 1947 (STA) which may be provided to local

authorities under legal gateways provided by legislation subsequent to the STA. Disclosure of information obtained under the STA (e.g. employment and earnings of employees), or statistics compiled wholly or partly from that information, is permitted by the Employment and Training Act 1973, as amended by the Employment Act 1988, to persons listed within section 4(3) of the 1973 Act. Some of these gateways are complex, take considerable time to set up, sometimes as long as two years, and are highly specific in terms of the data to be shared and the purpose for which sharing is required. Box 8 gives an example of the complexities of this process.

In their consideration of these difficulties, Thomas and Walport made the following recommendation:

Where there is a genuine case for removing or modifying an existing legal barrier to data sharing, a new statutory fast-track procedure should be created. Primary legislation should provide the Secretary of State, in precisely defined circumstances, with a power by Order, subject to the affirmative resolution procedure in both Houses, to remove or modify any legal barrier to data sharing by:

- *repealing or amending other primary legislation;*
- *changing any other rule of law (for example, the application of the common law of confidentiality to defined circumstances); or*
- *creating a new power to share information where that power is currently absent.*

(Thomas, R. and M. Walport 2008 'Data Sharing Review Report')

In their response to this recommendation, the Ministry of Justice stated:

The Data Sharing Review recognises the default position in the public sector has been to legislate, creating large numbers of specific legal gateways for sharing personal information. There are occasions when the requirement to share data should be put into primary legislation. Where this is evident, primary legislation should be sought as appropriate.

(Ministry of Justice 'Response to the Data Sharing Review Report' 24 November 2008)

The Taskforce has given careful consideration to the legal issues associated with linkage between and access to administrative data. While there are possibilities to overcome these issues in specific circumstances, the complexity of the situation will remain given the wide permutation of linkage requests that the research community will demand. The Taskforce has noted the difficulties that the last government faced in putting in place a broad and generic data sharing power via the Coroners and Justice Bill in 2009[9]. However, the Taskforce believes that the preferred approach, following up on the Walport-Thomas recommendations, should be to establish a **generic legal gateway** for research access to and linkage between administrative data. Such a gateway will be more efficient administratively, will be clear in the scope of its purpose, will allow research to be conducted and completed in timeframes which are commensurate with the needs of evidence-informed policy making and will provide a degree of consistency in the decisions made by data holders which is currently lacking. The safe and secure approach to access and linking proposed in this report provides the environment for legislative change.

[9] The clause inserted into the Coroners and Justice Bill, which did not survive Parliamentary scrutiny, related to the sharing of personal data. The recommendations of the Taskforce are concerned with access to and linkage between de-identified personal data.

3.4 A dual-track approach

Identification of the need for new legislation is an option that the Taskforce has considered carefully, given the pressures on the legislative timetable and the complexities associated with the responsibilities of the devolved administrations. While new legislation is strongly recommended, the Taskforce wants to ensure that the proposals it is making are not delayed or diffused as this process is undertaken. A dual-track approach recognises that much can be achieved while the legislative timetable takes shape. This includes setting up the ADRCs and the associated Information Gateway, establishing strong and robust governance arrangement, developing harmonised accreditation procedures, agreeing a common approach to consent for linkage and putting in place the resources required within departments and agencies to facilitate linkage. While new legislation can help unlock access to and linkage between data held by specific government departments, there is much to be achieved in advance of a new legislative environment which will improve and facilitate research access to administrative data whilst laying the foundation for a new legal gateway for data linkage.

3.5 Ethics and approval for research using administrative data

The Framework for Research Ethics, recently revised by the Economic and Social Research Council, outlines six key principles that underlie an ethical approach to research on human subjects. These are:

1. Research should be designed, reviewed and undertaken to ensure integrity, quality and transparency.
2. Research staff and participants must normally be informed fully about the purpose, methods and intended possible users of the research, what their participation in the research entails and what risks, if any, are involved.
3. The confidentiality of information supplied by research participants and the anonymity of respondents must be respected.
4. Research participants must take part voluntarily, free from any coercion.
5. Harm to research participants and researchers must be avoided in all instances.
6. The independence of research must be clear, and any conflicts of interest or partiality must be explicit.

Source: Framework for Research Ethics, Economic and Social Research Council, 2012

This framework, which applies to research funded by the ESRC, provides useful guidance for an ethical approach to research which makes use of administrative data from data subjects. Clearly, where information has already been collected from large numbers of data subjects and the expectation is that the data used for research will be de-identified, as is the case with most administrative data sources, the 1998 Data Protection Act always provides alternatives to consent as the basis for processing personal data legitimately. The view of the Information Commissioner's office is that consent is generally not needed to legitimise the conversion of personal data into a non-identifiable form. Indeed, the prospect of obtaining consent for data linkage would be prohibitively expensive and, even if it could be achieved, the biases such a procedure would introduce could invalidate the research process. It is for this reason that the recommendations for public engagement are made by the Taskforce (see section 5) in lieu of the second and fourth principles stated above.

The third principle, maintaining the confidentiality of the identity of persons whose data are to be used for research, is at heart of the proposal to use a trusted third party approach to data linkage and for access to be provided via a secure data access facility.

To ensure that the remaining principles are adhered to, all higher education institutions and research institutes have established procedures overseen by research ethics committees which allow them to monitor research proposals for ethical approval prior to any research being undertaken. Likewise, the major research infrastructures which house significant research data collections have their own ethics committees which safeguard the use of data against reputational damage or any physical or mental harm to data subjects.

The Taskforce recommends that all applications to use data held by the ADRCs should pass through an appropriate ethics committee to ensure that the confidentiality of data subjects is upheld and that no harm, including reputational damage, can come to data holders, data subjects, staff of the ADRCs, researchers and their institutions. Where the research application comes from an organisation which does not have an ethics committee, the ADRC should make special arrangements for such approval to be obtained from an appropriate ethics committee.

3.6 Consent for linking administrative data to survey data

The issue of consent from data subjects is appropriate where they have voluntarily given personal information in a survey. Linking administrative data to personal records obtained by survey techniques will normally require that the data subject should have given explicit permission for such linkage to take place. Such consent for linkage has, at times, been vague and unspecific. There is a need to review how different holders of administrative data have specified the wording of consent for linkage questions on various surveys, with the aim of developing a standard set of consent for linkage questions for specific types of administrative data.

The perception of privacy issues can also lead departments to design data collection processes that may limit re-use of data. For example, where data are collected by consent the confidentiality undertakings given to citizens may limit use of the data to the original department only. There is no standard consent wording used across government and no requirement that it should be written in a way that considers the possibility of re-use.

3.7 Recommendations

- Government departments, Local Authorities, agencies and other public bodies face different legal restrictions on the nature of the access they might provide for research using data they control. In such cases, specific legal gateways have been established to resolve this issue. However, recent experience demonstrates that current link-specific gateway legislation is both cumbersome and inefficient. The Taskforce recommends that primary legislation should be sought to provide a **generic legal gateway** for research and statistical purposes that enables efficient access to, and linkage between, administrative data held by different government departments, agencies and other statutory bodies.

- An agreed set of ethical standards should be produced, drawing on well-established ethical guidelines and covering the research uses to which administrative data (and administrative data linked to other types of data, including surveys) may and may not be put.

- In situations where linkage is proposed between large and de-identified datasets, consent for linkage is not required under the 1998 Data Protection Act. Where linkage involves the addition of administrative data to information collected by survey methods, it is both practicable and desirable to request consent for linkage from data subjects, even though the linked data will be de-identified prior to analysis. A common approach to the method of obtaining consent will be developed which will improve the efficiency of consent procedures and permit wider sharing of such linked data for research purposes.

Box 8

The difficulties of establishing and utilising a legal gateway

The Office for National Statistics (ONS) has recently been seeking access to selected demographic information as part of its work to look at possible alternatives to a traditional Census. The data relate to individuals who have interacted with the Social Security or Revenue systems and are held on the Customer Information System (CIS), which is administered by the Department for Work and Pensions (DWP). By reusing these data for statistical purposes ONS can potentially reduce its costs and respondent burden.

ONS stated its desire to access CIS data in October 2010 and the work to acquire these data for testing purposes started in March 2011. Once the variables required had been identified with DWP, the next stage was to find out how they could be legally accessed. In this case the ownership of the information was spread across three government departments and the legal position for each department had to be assessed. It soon became clear that new secondary legislation (a study-specific gateway) would be required in both Westminster and Belfast. Additionally, as the information in the system had been collected over a number of years and under different consent/fair processing notices, it also became evident that some variables could not be easily lawfully accessed and so needed to be dropped. Every variable to be accessed had to be specified and separately justified in the gateway request. While the legal position was being assessed, the Privacy Impact Assessment (PIA) and Business Impact Assessment (BIA) were carried out. It was decided that to support the PIA, interested parties and privacy groups should be invited to a workshop where any concerns they may have could be raised and addressed. The BIA identified the level of security the data would require and allowed ONS to start considering how best to manage the data once received.

Once this preliminary work had been completed, which took 4-5 months, the lawyers were in a position to start drafting the legislation. For this data share, there was additional complexity as, although the information required is held on one system, the CIS has never been defined in law; reference had to be made to the legislation that allows the data to be collected initially. Once drafted, it was a case of securing Parliamentary and Assembly time, the Westminster debates took place either side of the Easter recess with the Northern Ireland debate in June 2012. By mid-June 2012 a position had been reached whereby the data could be legally transferred. Then, before any information can be physically transferred, the data owners need to be assured that appropriate data security and management processes are in place. As the data will be used in conjunction with other datasets owned by other government departments, ONS has to ensure that all data owners are content with the arrangements being put into place. The current expectation is that the first data will be received by the end of November 2012, more than two years after ONS had stated its desire to access the data, and that it will only partially meet the original aims of the data share.

4 RESEARCHER ACCREDITATION AND TRAINING

4 Researcher accreditation and training

Research access to administrative data, especially in situations where data have been linked or are deemed to contain sensitive information, requires that the researcher fully understands the duty of care they have to preserve the confidentiality of the data to which they have access. Even though access will only ever be granted to de-identified personal data, there remain risks of disclosure which require researchers to understand the need for access in a safe setting and the requirement to conduct their research in ways which are consistent with the minimisation of such risks.

4.1 Current practice for accreditation

The Statistics and Registration Service Act 2007 established the legal status of an 'approved researcher', an individual to whom the Office for National Statistics (ONS) has granted access for statistical research to personal information held by the ONS. The Act provided further guidance to the ONS in establishing the criteria by which it will determine whether to grant access, requiring that access only be granted to a 'fit and proper person' and that the purpose for which access is to be granted should be stated. Building upon procedures originally introduced to authorise access by researchers to data held in its Virtual Microdata Laboratory, the ONS has established rigorous procedures which can confer approved research status with a minimum delay[10].

In line with these developments, the Economic and Social Research Council (ESRC) introduced the concept of an 'accredited researcher', developed to facilitate access by researchers to data held within the ESRC Secure Data Service. This process is similar to that operated by the ONS for approved researcher status, but is designed for access to datasets containing sensitive personal information held within the Secure Data Service which are not designated as national statistics.

4.2 A UK-wide system of accreditation

Some government departments have set up similar procedures of their own, but these are not consistent and appear to be causing some confusion over their interpretation. There is a clear need for one UK-wide system of accreditation to be adopted. This could build upon the collaborative approach adopted by the ONS and the ESRC Secure Data Service, adapting this where necessary to accommodate any essential additional requirements for research access to specific administrative datasets. It should also have regard for developments internationally in researcher accreditation, seeking to ensure that the standards that are set are consistent with international efforts to widen cross-national access to administrative data[11].

> "THERE IS A CLEAR NEED FOR ONE UK-WIDE SYSTEM OF ACCREDITATION TO BE ADOPTED."

[10] See http://www.ons.gov.uk/ons/about-ons/who-we-are/services/virtual-microdata-laboratory/accessing-the-vml/how-to-access-the-vml/index.html

[11] See, for example, the work being conducted within *Data without Boundaries* (http://www.dwbproject.org/)

4.3 Training researchers – data analysis and data security

Accreditation requires researchers to conduct their research in ways that are consistent with the lowest possible risk of inadvertent disclosure of the identities of individuals. This requires that the researcher understands the nature of these risks and is trained to carry out their research in ways which minimise disclosure. Both the ONS Virtual Microdata Laboratory and the ESRC Secure Data Service require approved or accredited researchers to have undertaken training in disclosure control and in the manner in which they should conduct their research in a secure data access facility. The Taskforce regards such training as an essential component of data access procedures and that this should be regarded as an essential pre-condition of research access to administrative data. To guard against situations where training in access procedures becomes out-dated, refresher course should also be required on a regular basis.

In line with the penalties for breach of conditions of access within the Statistics and Registration Service Act 2007, and data protection legislation, it is proposed that accreditation will be accompanied by a set of sanctions to be applied in a proportionate manner to researchers and/or their institutions in the event of any breach of the conditions of access to administrative data.

4.4 Recommendations

- All researchers wishing to analyse administrative data through the ADRCs should be required to gain accreditation status.
- **A single accreditation process** needs to be designed which builds on best national and international practice and is acceptable to all UK holders of administrative data.
- This accreditation process will require short course training including: best practice methods for working in a secure data environment; legal and ethical issues related to the use of individual records; methods of disclosure control; and an introduction to the variety of types of routinely collected administrative data.
- To retain accreditation status, researchers should be required to undertake an online update course on a regular basis.
- Where there is no recourse to legal penalties for any breach of data access conditions, accreditation will be accompanied by a set of sanctions to be applied in a proportionate manner to researchers and/or their institutions in the event of any such breach of the conditions of access to administrative data.

5 A STRATEGY FOR ENGAGING WITH THE PUBLIC

5 A strategy for engaging with the public

Public support for and confidence in the research use of administrative data can only develop if there is clarity about the purpose of the research, the data to be processed, and the safeguards that will be in place to prevent any abuse of such data. This requires that the research aims, the processes of governance for the use of data and data management plans are transparent and accessible to a wide audience, necessitating engagement with the public (and public opinion formers) which is as yet not a routine feature of most research. However, it is also important to be clear about the purposes of public engagement in this context.

5.1 What is public engagement in research?

Public engagement refers to a range of participatory activities through which the citizen may be involved with the research process. Drawing on a review of existing classifications of engagement, Aitken (2010) offers a straightforward tripartite division of engagement into: awareness-raising, consultation, and empowerment activities. The purpose of these types of activities, the outcomes that are desired and the potential methods by which they can be achieved is summarised in Box 9.

5.2 Administrative data and public engagement – what evidence do we have?

Most of the evidence accumulated to date on public engagement relates specifically to public attitudes and awareness of linkage and/or secondary uses of medical data. This may in part be due to the high profile of medical research and also because of the extent of existing linkage studies. A number of studies have indicated general public support for uses of medical data in health research[12], although this support is often tempered by concerns about confidentiality and individual privacy[13]. For example, Heath (2010: 87) has noted that: 'Tension exists between consumers' expectations of individual privacy and recognition of the utilitarian gains available through secondary uses of medical data for health research'.

Box 10 gives examples of recent initiatives that have sought to provide information to assist with the development of research access to and linkage between various types of data.

Box 9

Nature of Engagement	Purpose	Desired Outcome	Potential Methods
Awareness-raising	Information provision and public education about the need for research based on administrative data.	Greater public acceptance or legitimacy for use of administrative data.	Media campaigns. Public exhibition/ presentations. Leaflets.
Consultation	To gain insight into public opinion/ views on use of administrative data for research.	Creation of appropriate/socially acceptable research usage of administrative data.	Surveys. Focus groups.
Empowerment	To work with the public enabling them to play key roles in decision-making regarding the administrative data to be accessed/linked.	Greater social capital. Capacity building. Enhanced democracy.	User panels. Citizens' juries. Membership of the Governing Board.

[12] Haddow *et al.* 2007, Kass *et al.* 2003, Stone et al. 2003
[13] Damschroder *et al.* 2007, Page and Mitchell 2006, Willison *et al.* 2007

5.3 What does the evidence reveal about public attitudes to administrative data access and linkage?

There is broad, though not unconditional, support for uses of administrative data for research. However, several studies have suggested that it may be important to demonstrate clearly the value of such uses of data. There is evidence that public support may be contingent on data access or linkage being demonstrated as being of public benefit.

A key theme is trust. Public support is higher when individuals trust the person or organisation accessing/handling their data. Trust may also be related to transparency, in that where the purposes of data use are clearly and openly communicated there may be greater support. For example, a Scottish Government study of public attitudes towards linkage of health, social care and housing support data found that people 'would like further information about how their data are accessed and by whom. In particular, they would like to be informed of the purpose and/or outcome of studies' (Aitken 2011: 13).

The literature has also indicated that consent, the need for individuals to feel they have some control of their data and its use, and anonymisation are important considerations shaping public responses. However, in general the emphases on individuals' control and respect for confidentiality are balanced by acknowledgements of the importance of enabling research to go ahead in the public interest. As such, whilst it is clearly important that individuals' confidentiality and autonomy are respected, members of the public appear to recognise the importance and relevance of data use for research.

5.4 Recommendations

- The Taskforce recognises the need to ensure that members of the public are aware of these developments in access to administrative data for research purposes, and that public concerns are given due attention. A strategy should be developed which will provide readily accessible information about the benefits of improved access to and linking between administrative data, and the measures being enacted to minimise risks of disclosure and to prevent inappropriate use of such data.

The strategy should encompass procedures for raising public awareness about the need for research based on administrative data, the nature of public consultation and the roles that members of the public could play in decision-making regarding the administrative data to be accessed and linked.

- To help develop this strategy, the Governing Board, which will include lay members, should steer the plans for public engagement prepared by each of the Administrative Data Research Centres, ensuring that there is no unnecessary duplication of efforts and suggesting ways in which these activities associated with these plans may be undertaken in the most efficient and effective manner.

- To assist with its implementation the Taskforce recommends that each of the proposed ADRCs should appoint a **public engagement and communications officer**, to lead engagement with a wide audience and promoting dialogue about the research benefits that could accrue, the safeguards that are required to prevent any misuse of data and how both could be effectively and collaboratively achieved.

- The ADRCs should produce plans for public engagement. They should collaborate to plan and hold public events to explain the work they are doing and to generate debate about the academic and broader social and economic benefits that derive from research using administrative data and the measures taken to ensure that the identities of individuals cannot be revealed.

Box 10: Examples of recent public engagement initiatives

Scotland-wide Data Linkage Framework for Statistics and Research

To supplement the written consultation, *A Scotland-wide Data Linkage Framework for Statistics and Research*, the Scottish Government commissioned Ipsos MORI, along with Professor Sarah Cunningham-Burley and Dr Claudia Pagliari from the Centre for Population Health Sciences at the University of Edinburgh, to undertake a series of public deliberative events. The overall aim of the events was to explore the acceptability of linking personal data for statistical and research purposes, thereby identifying particular sensitivities and exploring mechanisms for overcoming concerns. All participants in these events recognised potential benefits of data linkage but most also had concerns about it. Participants often assumed that linked data would include personal identifiers. When they were told that this would generally not be the case, many became more comfortable with the idea. However, a minority contended that anonymised data could always be linked back to personal identifiers by anyone with the necessary knowhow. A number of safeguards were identified to maximise public confidence in the framework, including that an oversight body should be established comprising highly qualified professionals and lay members, with responsibility for granting or refusing linkage requests.

Public Acceptability Research Report

In February 2012 the Office for National Statistics 'Beyond 2011' Programme commissioned a study seeking the public's views on census-taking and the sharing and holding of personal information for statistical purposes. The research was designed to assess public understanding and gather information on views about how government collects, holds and shares socio-demographic information. The study revealed that public benefits of data sharing are not widely understood; while more than half of the public thinking that data sharing could lead to improved data quality, 20 per cent do not demonstrate any understanding of the potential benefits. It also showed that support for the holding of personal information in a database used for statistical purposes is high, with well over half of the public supporting the concept and 19 per cent neither agreeing nor disagreeing. A large majority believe that personal details such as name, address, date of birth and sex should be held on a central database and 45 per cent think that it would reduce government costs and help to deliver services more effectively.

Office for National Statistics 'Beyond 2011' programme, *Stakeholder Engagement and Communication Project* (June 2012)

6 RESOURCE COSTS

6 Resource costs

The Administrative Data Taskforce has been supported and driven by a range of social science and medical research funders, some of which are represented on the Taskforce. As the recommendations resulting from the Taskforce will not only benefit the academic community, but the wider researcher and policy making community, a significant number of government departments and the devolved administrations have participated as members of the Taskforce and in the development of these recommendations. A broad consortium of funders drawn from research funders and government departments will consider their implementation. The anticipated resources required to fund the recommendations set out in this Report for a period of five years is **£30 million**. This funding will support the interlinked components of the recommendations that represent a mix of 'one-off' investment and recurrent costs over an initial five-year period.

6.1 Administrative Data Research Centres (ADRCs)

The Taskforce recommends that four national Administrative Data Research Centres should be established, giving due regard to the needs of the devolved administrations and developments therein. While the detailed organisation and structure of each centre may reflect national variations in access arrangements, common features will consist of:

- a leadership team including a Director, researchers, technical, and statistical staff, together with public engagement and secretariat support;
- a programme of research aimed at addressing important research questions and improving understanding of the research value and quality of administrative data;
- a secure access facility in each ADRC which meets the most stringent international standards, with powerful analytical software, with free access to eligible researchers;
- commissioning and undertaking linkage of data where personal identifying information is not held in the ADRC, but is matched through a third party service;

- monitoring systems and expertise to ensure that proposals for access to and linkage between administrative datasets are legal, viable, technically feasible and that the research skills necessary to conduct the research efficiently are available;
- an Information Gateway, likely to be managed by one ADRC, to be the first port of call for prospective researchers, will be used to manage applications to use data, providing information about available datasets, and guidance about the knowledge and credentials that are needed to gain access to and use a wide variety of types of administrative data for their research.

Each ADRC will have its own staff and premises, and may build upon, take advantage of, and extend existing structures and expertise. Work on data linkage has reached different stages in the four nations, and ADRCs will potentially build on relevant initiatives where they exist.

ADRCs will also need to maintain an audit trail of all activity relating to data access and linking, and operate output control systems. They must also develop public engagement strategies and activities, led by a public engagement and communications officer in each ADRC. The Taskforce recognises that the level of funding for each ADRC will vary across the countries of the UK depending on the base infrastructure in place and variation in the anticipated volume of linkage requests, and so the resourcing will reflect that base position. The initial resourcing of **£23 million** for the ADRCs and their associated activities, including a programme of research, will provide baseline support for the network of ADRCs for an initial five-year period. Further resourcing, if available, would expand the impact and usage of the ADRCs and will be required for activities beyond the initial five years.

6.2 Establish a UK Governing Board

The Taskforce recommends the establishment of a UK-wide Governing Board for the ADRCs to perform key strategic and operational functions set out within this Report.

The Governing Board would have an independent chair and secretariat, and include senior representatives from the relevant government departments, agencies, and devolved administrative bodies that provide access to their data through one or more ADRC. Funders, national and international experts, lay members, and the directors of the ADRCs will also be included in the membership. It is anticipated that the total membership will consist of 20 to 25 persons. Sub groups of the Governing Board will be established to support requests for access to and linkage between administrative datasets, as well as linkage to non-administrative datasets. The Governing Board is also expected to commission work and co-opt members if necessary to provide expertise in areas relating to ethical standards, international experience, linkage methodology, security, legal advice and other issues.

The establishment and running of the UK Governing Board itself and its sub groups for the initial period of five years, together with the envisaged activity to commission a number of expert reports as well as legal advice, suggests the resources required total **£1 million** for the first five years.

6.3 Support the development of the UK-wide researcher accreditation process and the provision of associated training courses

The Taskforce recommends that a UK-wide researcher accreditation process be implemented, along with associated training courses. Training will be available for researchers in the use of safe settings for data access, data disclosure, and secure data management, with an annual online update course for all accredited researchers.

The implementation of this recommendation will build on and extend existing expertise and training provided within government and by the ESRC-funded Administrative Data Liaison Service (ADLS). Resourcing of **£1.5 million** for the first five years is initially required to develop and roll out this training and accreditation, although further resources may be required as the user base expands.

“TRAINING WILL BE AVAILABLE FOR RESEARCHERS IN THE USE OF SAFE SETTINGS FOR DATA ACCESS, DATA DISCLOSURE, AND SECURE DATA MANAGEMENT, WITH AN ANNUAL ONLINE UPDATE COURSE FOR ALL ACCREDITED RESEARCHERS.”

6.4 Support government departments in enhancing and providing data to ADRCs

The Taskforce recognise that, in order to enable government departments to support data retrieval, the creation of appropriate metadata, and the agreement on and implementation of agreed standards in data management, support for key activities at the interface between the relevant government departments and the UK Administrative Data Research Network would be required. These activities might include: undertaking the data extraction and preparing relevant meta-data; providing ongoing support for studies requiring access to their data; and contributing to the governance of the ADRCs. This partnership approach may also include joint research programmes.

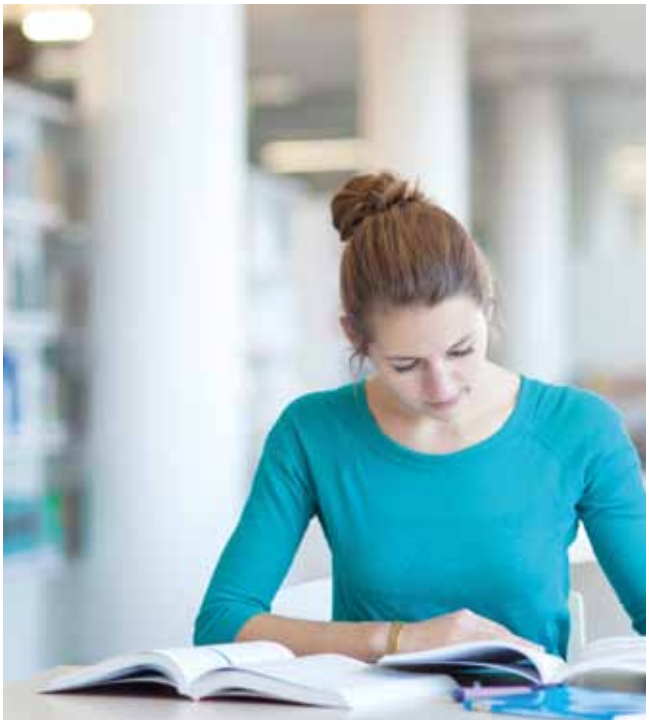
A funding level of **£3 million** over the five years to be supplemented by funding from government (in-kind or cash) would provide the initial support for key government departments to participate in the implementation of these recommendations.

6.5 Assist Higher Education Institutions with the installation of secure rooms and the necessary equipment to provide remote access to ADRCs

Researchers will be able to gain access to data in a secure environment in a secure setting within an ADRC, or via submitted statistical syntax to the expert team in an ADRC. However the most efficient research environment for many researchers who are not based close to one of the safe settings in an ADRC is for remote access via a secure 'thin' client link to an ADRC. This method of access is in place already via the ONS Virtual Microdata Laboratory (VML) and also the ESRC-funded Secure Data Service. It requires a secure room in an approved institution to be set up, with appropriate security and monitoring at a level equivalent to that for the planned ADRCs. Most higher education institutions do not currently have this facility, and will require funding for refurbishment, technical support, appropriate equipment, security arrangements, as well as staffing costs (research conducted in the remote sites will need to be overseen). Resourcing for this infrastructure will need to increase as the number of secure rooms increase. However an initial **£1.5 million** investment will establish and support a network of up to ten secure facilities in the first instance.

6.6 Summary of resources required

The table opposite shows the estimated cost of establishing and running the UK Administrative Data Resources Network over a five year period. While these estimates are based upon the best available current information, they may be subject to revision as detailed plans for the network emerge.



Areas	Resourcing over 5 years
Funding for ADRCs	£23 million
Funding for UK Governing Board	£1.0 million
Funding for accreditation and training	£1.5 million
Funding for government departments	£3.0 million
Funding for remote secure access facilities	£1.5 million
TOTAL	**£30 million**

7 SUMMARY: MOVING FROM RECOMMENDATIONS TO ACTIONS

7 Summary: moving from recommendations to actions

The recommendations of the Taskforce are summarised in the executive summary and elaborated in more detail in the subsequent sections. This concluding section provides an outline of the actions that are needed to implement these recommendations. For each set we present a timeline for the activities concerned and identify the lead responsibility. The timings shown are effective from the beginning of the next financial year, April 2013.

An Administrative Data Research Centre (ADRC) should be established in each of the four countries in the UK

Action	Timeline from April 2013	Lead Responsibility
Establish the UK-wide Governing Board	Within two months	ESRC together with the reporting authority for the Governing Board
Commission the ADRCs	Release the specification for the ADRCs within two months and launch within 12 months	ESRC, in consultation with other funders
Commission the Information Gateway	Release the specification for the Information Gateway within two months and launch within 12 months	ESRC, in consultation with other funders

Legislation should be enacted to facilitate research access to administrative data and to allow data linkage between departments to take place more efficiently

Action	Timeline from April 2013	Lead Responsibility
Explore requirements of proposed legislation and implications for existing legislation	Ongoing	Government, liaising with relevant members of the Governing Board
Identify appropriate Bill to enable legislative change	Ongoing	Government
Develop agreed set of ethical principles	Within ten months	Governing Board

A single UK-wide researcher accreditation process should be established

Action	Timeline from April 2013	Lead Responsibility
Establish a UK-wide researcher accreditation process	Release the specification for the accreditation process within two months and launch within 12 months	ESRC, liaising with ONS and relevant government departments
Establish an online accreditation update course	Launched within 24 months	Accreditation process award winners
Establish training courses	Release the specification for the training courses within two months and launch within 12 months	ESRC

A strategy for engaging with the public should be instituted

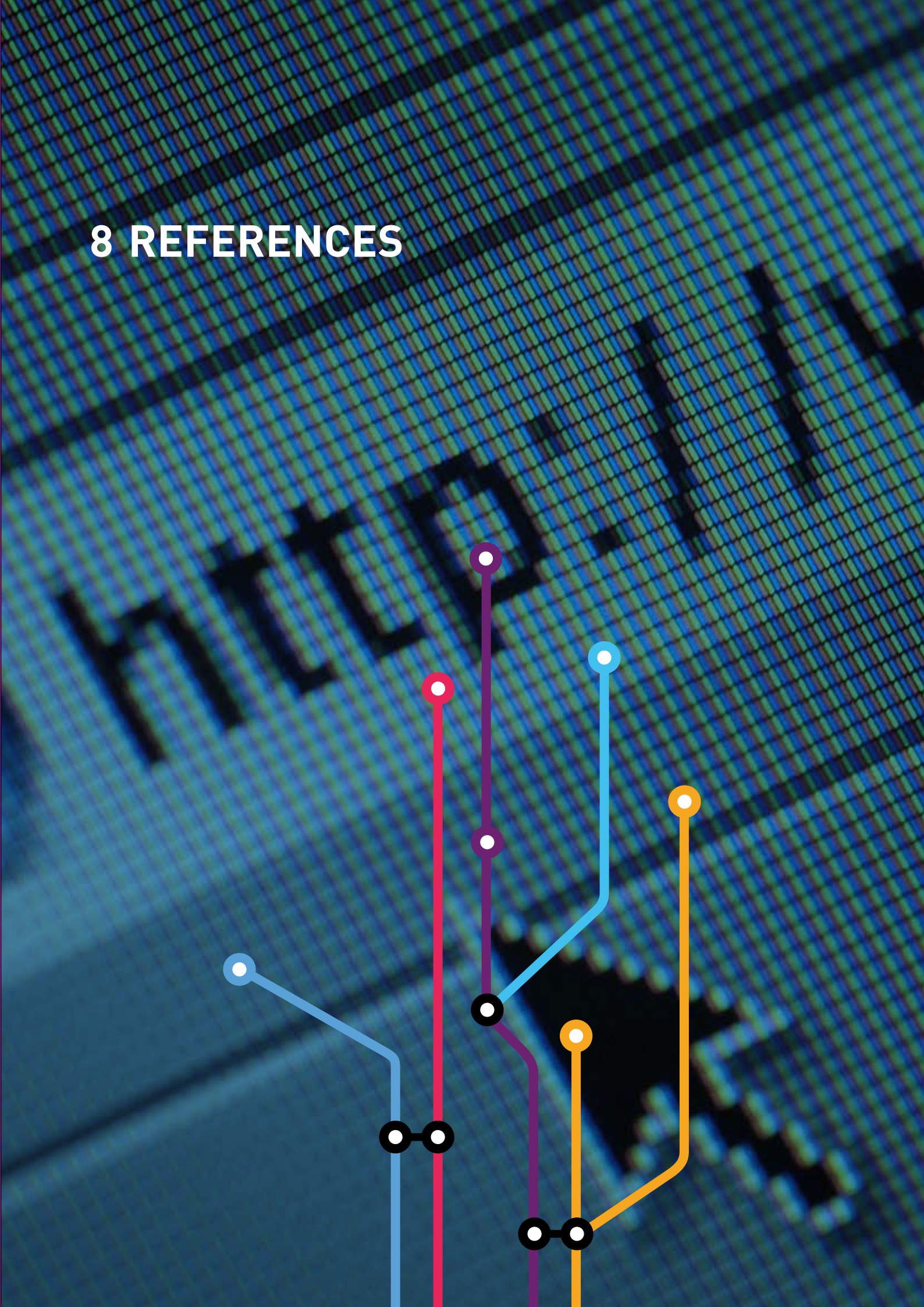
Action	Timeline from April 2013	Lead Responsibility
Commission public engagement review	Within six months	Governing Board
Public engagement and communications officers appointed	Within 18 months	ADRCs
Public engagement strategies and activities embedded within ADRCs	Within 24 months	ADRCs and Governing Board

Sufficient funds should be put in place to support improved research access to and linkage between administrative data

Action	Timeline from April 2013	Lead Responsibility
Negotiate programme funding lines	Within three months	Funders, Governing Board and government departments
Establish and support a network of remote secure access facilities in HEIs	Release the specification for the network of the remote secure access facilities within two months and launch within 12 months	ESRC

The work required to put these recommendations into effect will involve much detailed planning and collaboration between those who will be funding the activities, those responsible for establishing the governance arrangements and the holders of administrative datasets. The Taskforce firmly believes that the resources required, together with mechanisms proposed to improve access to and linkage between de-identified administrative data, will stimulate research use of these valuable data and will enhance the evidence base for better policy making. In their entirety, these actions will form the most significant step forward that has ever been taken across the UK to improve access to and linkage across government administrative data for research and policy purposes. The **UK Administrative Data Research Network** will help realise the potential we have to use these resources in ways that are efficient, innovative and informative.

8 REFERENCES



8 References

Aitken, M. (2011). *Linking Social Care, Housing and Health Data: Social Care Clients' and Patients' Views*. Scottish Government Social Research.

Cabinet Office (2012). *Unleashing the Potential Open Data White Paper* June 2012 Cm 8353.

Damschroder, L.J., J.L Pritts, M.A. Neblo, R.J. Kalarickal, J.W Creswell and R.A. Hayward (2007). 'Patients, privacy and trust: Patients' willingness to allow researchers to access their medical records' *Social Science and Medicine* 64: 223-235.

Economic and Social Research Council (2012) *Framework for Research Ethics* (FRE) 2010 (Updated September 2012). See http://www.esrc.ac.uk/about-esrc/information/research-ethics.aspx

Haddow, G., G. Laurie, S. Cunningham-Burley and K.G. Hunter (2007). 'Tackling Community Concerns about Commercialisation and Genetic Research: A modest interdisciplinary proposal' *Social Science and Medicine* 64: 272 – 282.

Heath, J. (2010). 'Emerging Consumers View of Secondary Uses of Medical Data' *IEEE International Symposium on Technology and Society*.

Information Commissioner's Office (2012). *Anonymisation: managing data protection code of practice.* See http://www.ico.gov.uk/news/latest_news/2012/~/media/documents/library/Data_Protection/Practical_application/anonymisation_code.ashx

Kass, N.E., M.R. Natowicz, S. Chandros Hull, R.R. Faden, L. Platinga, L.O. Gostin and J. Slutsman (2003). 'The Use of Medical Records in Research: What do patients want?' *Journal of Law, Medicine and Ethics* 31: 429-433

Office for National Statistics (2012). *Public Acceptability Research Report Stakeholder Engagement and Communication Project* Beyond 2011 Programme.

Page, S.A. and I. Mitchell (2006). 'Patients' opinions of privacy, consent and the disclosure of health information for medical research' *Chronic Diseases in Canada* 27 (2): 60-67.

Stone, M.A., S.A. Redsell, J.T. Ling, and A.D. Hay (2005). 'Sharing Patient Data: Competing demands of privacy, trust and research in primary care' *British Journal of General Practice*.

Thomas, R. and M. Walport (2008) *Data Sharing Review Report prepared by for the Ministry of Justice*. See: http://www.justice.gov.uk/docs/data-sharing-review-report.pdf

Willison, D.J., L. Schwartz, J. Abelson, C. Charles, M. Swinton, D. Northcup, and L. Thabane (2007). 'Alternatives to Project-specific Consent for Access to Personal Information for Health Research: What is the Opinion of the Canadian Public?' *Journal of the American Medical Informatics Association* 14 (6): 706-712.

Appendix 1 Terms of reference and membership of the Taskforce

Terms of Reference:

The key role of the Taskforce will be:

- identification of potential risks and benefits from increased research use of administrative data;
- identification of likely resource implications arising from increased research use of administrative data;
- the development and introduction of common procedures to provide more efficient access to administrative datasets;
- clarification of the legal situation governing the use of routine data;
- clarification of when consent is required and what consent procedures should be used;
- identification of possible need for legislative change to improve access to administrative data for research.

Other priorities for the Taskforce include:

- the development of agreed methods for data linkage;
- further development of a 'metadata authority' to assemble and disseminate information relating to the use of administrative data for research, for details of data linkage procedures and for the preservation of information relating to the quality of various administrative datasets as research resources;
- agreement regarding the potential preservation and reuse of linked data;
- procedures to raise public awareness of the benefits arising from research use of administrative data and data linkage;
- guidance on data access, including the use of secure data access facilities and how the research environment should be controlled;
- proposals for how the quality of administrative data may be assessed.

Timetable and deliverables

The Taskforce will deliver its recommendations to Ministers within no more than 12 months of its formation.

The recommendations will be presented in a publicly available Report together with an agreed Action Plan, endorsed by Members of the Taskforce.

There should be agreement on how the implementation of the recommendations and Action Plan will be monitored. The Report may wish to propose that new or existing body take responsibility for oversight of the implementation of the recommendations.

Membership

Membership will have between 15 and 20 members drawn from relevant government departments and agencies, funders of research in the medical and social sciences, representatives from the research communities, the private sector, and the Office for National Statistics.

Chair	Secretariat
Sir Alan Langlands Higher Education Funding Council for England (HEFCE)	**Vanessa Cuthill** Economic and Social Research Council (ESRC)

Members	
Gill Aitken Department for Work and Pensions (DWP)	**Guy Goodwin** Office for National Statistics (ONS)
Iain Bell (to July 2012) / Rebecca Endean Ministry of Justice (MoJ)	**Roger Halliday** Scottish Government
Iain Bourne The Information Commissioner's Office (ICO)	**Mike Hawkins** HM Revenue and Customs (HMRC)
Paul Boyle Economic and Social Research Council (ESRC)	**Tim Kelsey (to April 2012) / Charlotte Alldritt** Cabinet Office
Heather Brown Department for Business, Innovation, and Skills (BIS)	**Peter Knight** Department of Health
Kate Chamberlain Welsh Government	**David Lynn** The Wellcome Trust
Keith Dugmore Demographics User Group	**David Marshall** Northern Ireland Statistics and Research Agency (NISRA)
Peter Elias ESRC Strategic Advisor for Data Resources	**Stuart Sarson** Government Office for Science
Wendy Ewart Medical Research Council (MRC)	**Andy Sutherland** Health and Social Care Information Centre
David Frazer Department for Work and Pensions (DWP)	**Carole Willis** Department for Education (DfE)

Task force meetings were held on:

8 December 2011; 16 January 2012; 28 March 2012; 25 April 2012; 30 May 2012; 9 July 2012; and 11 October 2012

Appendix 2 Models for administrative data access and linkage

This appendix presents a review of the main models being used for access to and linkage between administrative datasets in the UK, summarising their strengths and weaknesses.

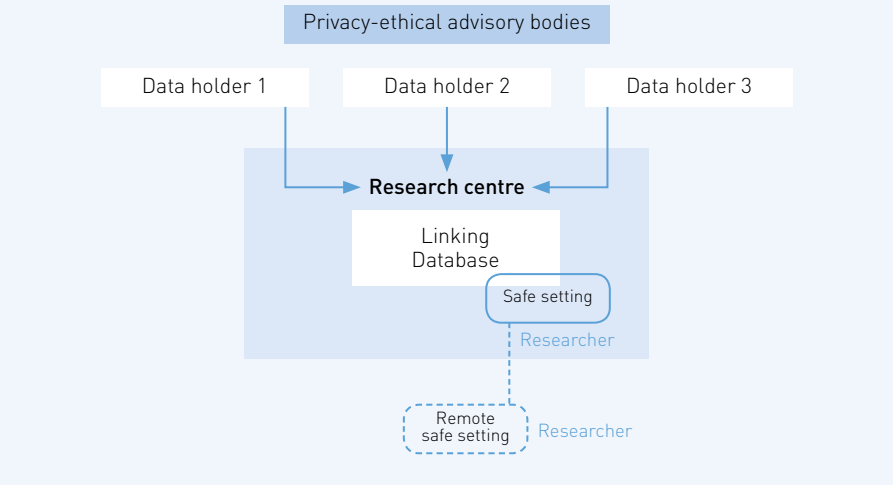
Model 1 – the single centre

Model 1 involves the transfer of data to a trusted research centre. The research centre then constructs a research dataset by linking data from different organisations and makes the resulting data available within a safe setting to authorised researchers. The data made available to researchers will almost always be reduced in content to ensure that within the controlled environment of the safe setting that the data is non-identifiable. Part of the control operated through the safe setting is over the level of information that can be removed. This will typically be only information that is entirely non-disclosive (i.e. is definitely non-identifiable). The data transferred will need to involve sufficient data to allow linkage (i.e. either an individual reference or uniquely identifying characteristics of the individual). This model is therefore highly likely to involve the transfer of identifiable data and therefore will represent a share (processing) of personal data.

Strengths

- Involves a limited amount of data transfer – reduces inherent danger of loss in transit
- Involves only one organisation – making scrutiny by overseeing bodies easier
- Involves only one organisation – more efficient e.g. fewer scheduling issues
- Processing by a single organisation – allows for the most effective calculation of linkage quality and assessment of bias

Model 1 – the single centre



- Research centre overtime can collect and organise metadata, collect programming and foster a deep understanding of administrative datasets

Weaknesses

- No 'visible' barriers to dishonest behaviour
- No structural aids (e.g. separation of functions) to support an argument that the processing is of non-personal data – i.e. that individuals cannot be identified
- Relies on the research centre being entirely honest

Model 2 – firewall single centre

Model 2 requires a physical separation of functions within the dataset construction process. Here the process of creating a link between records for the same individual and then the linking of the full research dataset are separated. In the first instance the data holders send a limited amount of information to a part of the research centre to enable matching between the different datasets. This would be limited to **only** the data necessary to achieve a satisfactory level of matching. This would often include name or parts of, date of birth, address information and an identifier unique and interpretable **only** by the data holder.

The indexer within the research centre then either deterministically or probabilistically matches records giving ones believed to be for the same person a unique **study identifier**. The study identifier and the relevant data holder identifier are then passed back to the data holders. Each data holder then adds the study identifier to their datasets removes any other identifier information and sends the research relevant data to a **separate** part of the research centre. The research management part of the research centre then links the various dataset using the study identifier and makes it available to researchers under the model 1 conditions. There is as a result a clear separation of function that acts to ensure that the identifier information cannot be used in conjunction with the research dataset to re-attached identifiers.

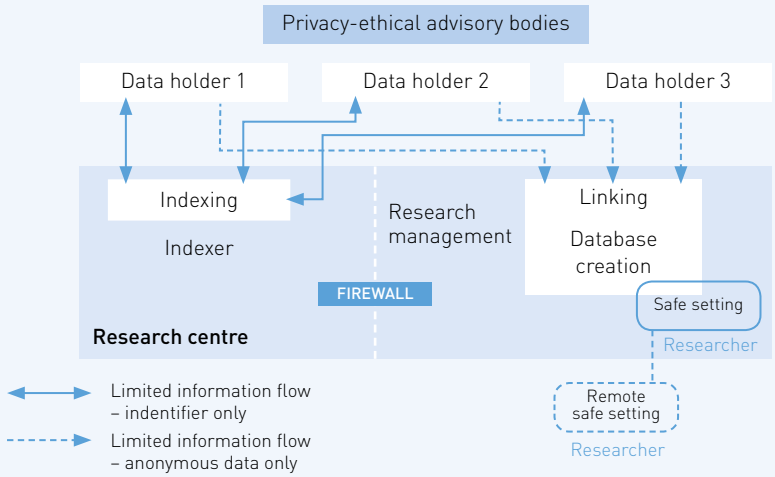
Strengths

- Offers a structural barrier to dishonest behaviour
- Involves a single organisation – making scrutiny by overseeing bodies easier
- Involves a single organisation – more efficient
- Processing by a single organisation – allows for the most effective calculation of linkage quality and assessment of bias although the sharing of information will be limited by the firewall
- Research centre over-time can collect and organise metadata, collect programming and overtime foster a deep understanding of administrative datasets

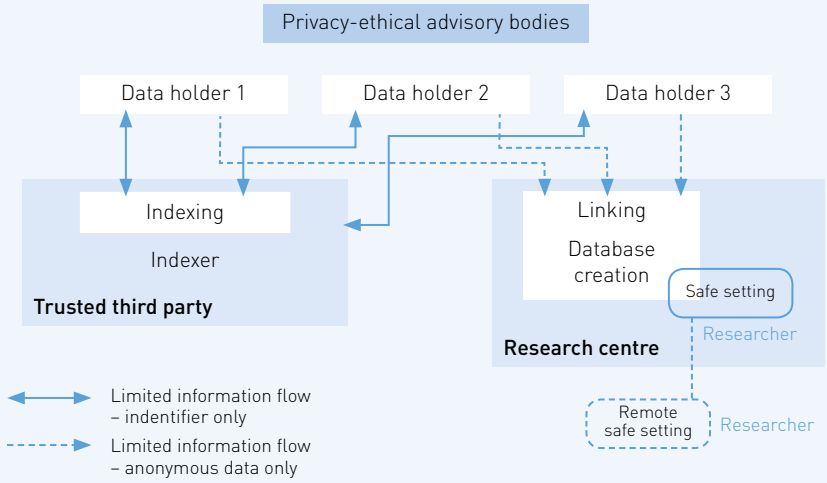
Weaknesses

- Though offering a barrier to dishonest behaviour – this is not particularly visible from the outside
- Relies on the research centre behaving honestly – i.e. adhering strictly to the firewall
- More data transfer – increases inherent danger of loss in transit

Model 2 – firewall single centre



Model 3 – trusted third party indexing



Model 3 – trusted third party indexing

Model 3 is a slight but important refinement of model 2. Instead of the indexing process taking place in a firewall separated part of the same organisation; it takes place in an entirely separate organisation. The Trusted Third Party (TTP) that will carry out the indexing importantly does **not transfer** any data to the research centre. Indeed any form of direct communication between the indexer and the research centre can be limited. This then means that the research centre does not need to hold identifier information (i.e. names, addresses etc.). Indeed it is possible that the research centre only needs to receive anonymous data that through restrictions on variables may also be non-identifiable (particular within the context of a controlled environment) and therefore is **not personal data.**

Strengths

- Offers a visible structural barrier to dishonest behaviour
- Structural aids (e.g. separation of functions) to support an argument that the processing is of non-personal data – i.e. that individuals cannot be (re) identified
- Research centre overtime can collect and organise metadata, collect programming and overtime foster a deep understanding of administrative datasets

Weaknesses

- Greater number of organisations involved – making scrutiny by overseeing bodies harder
- Greater number of organisations involved – less efficient
- More data transfer – increases inherent danger of loss in transit
- Processing by two organisations – barrier to effective calculation of linkage quality and assessment of bias

Model 4 – secure multi-party computation

Model 4 is an entirely different approach. Here there is no research centre but instead the researcher makes an approach to a set of data holders requesting that they set about a process of 'secure multi-party computation' at the end of which the relevant summary statistics are given to the researcher. The data holders do need to have a common identifier across the datasets to be linked and so a trusted third party may have to be used for this purpose. In summary secure multi-party computation involves the transfer of large quantities of data that allow the computation of the necessary statistics without revealing the values of the underlying data. For example in the case of secure regression, in the context of a simple vertically partitioned data set, where each data holder holds different sets of records on **all** of the individuals in any of the other datasets, data agencies need to exchange privacy-preserving matrix representing the structure of data but not its values between each other. This then allows through a chain process, that may have to be iterated, the construction, over all agencies, of the off diagonal elements of the relevant variance-covariance matrix across all the variables in the analysis. Each data holding body will also have to share the 'on-diagonal' elements of the variance-covariance matrix (i.e. the variances of each variable) allowing its complete construction.

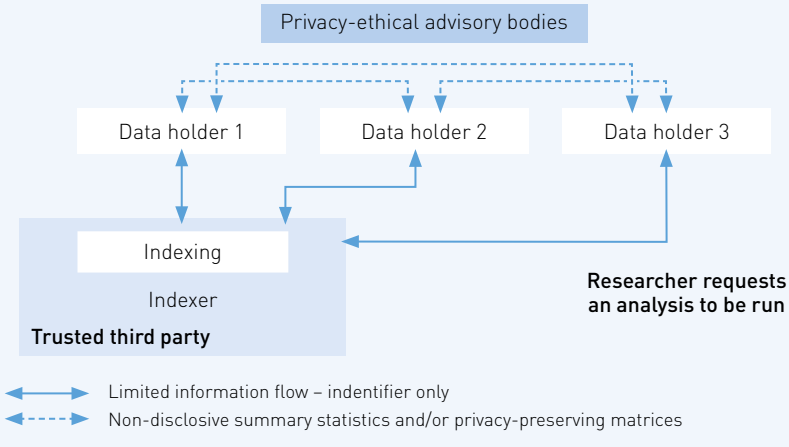
Strengths

- Offers a visible structural barrier to dishonest behaviour by researchers – they have no access to individual level data
- Arguably there is no transfer of personal data – transfer only of privacy preserving matrices and the types of analysis can be controlled so output will not be disclosive

Weaknesses

- A developing area of computer/ statistical science – many unknowns
- Some analysis are not as yet possible – may not be possible
- Involves a high degree of involvement of multiple dataholders in every analysis – automation may be possible but may not be desirable
- Requires the dataholders to provide research ready data (e.g. clean, robust, accurate etc.)
- Involves very large amounts of data transfer – however this will be in a privacy preserving form – loss/ interception should not therefore lead to disclosure
- Involves operations on very large matrices (n*n) – may require significant computing power to be available within all data holding organisations
- Much less opportunity to collect and organise metadata, collect programming and to overtime foster a deep understanding of administrative datasets

Model 4 – secure multi-party computation



Appendix 3 Definition of key terms

Accredited Researcher
Researcher who proposes to gain access to secure administrative data micro records and/or initiates a data linkage in a safe setting in order to answer a research question and has been approved as such via an accreditation process.

Administrative Data
Administrative data refers to 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.

Administrative Data Research Centre
An organisation, incorporating a secure data access facility, whose responsibility is the safe and legal provision of research datasets which incorporate national administrative data and the support of research using such data.

Anonymised Data
Such data have all personal identifiers removed and cannot be connected to the original person record. Anonymised data are suitable when no contact is needed with the participant or where the data do not need to be linked to any other data sources.

Confidential Information
Any information obtained by a person on the understanding that they will not disclose it to others, or obtained in circumstances where it is expected that they will not disclose it. For example, the law assumes that whenever people give personal information to health professionals caring for them, it is confidential as long as it remains personally identifiable. See also 'Personal data'.

Consent
Explicit agreement by the data subject for his or her personal data to be processed for a specific purpose.

Data Holder (data controller)
The person who decides the purposes for which, and the manner in which, personal data is to be processed. This may be an individual or an organisation registered with the Information Commissioner's Office.

Database
A collection of research data that is organised and allows its contents to be easily accessed, managed and updated. The type of database used depends on the requirements of the study. A common type is the relational database, where data are related to each other in a systematic manner so that they can be reorganised and accessed in a number of different ways. A database may house one or many datasets.

Data Linkage / Record Linkage
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. See also 'Indexer'.

Data Subject
An individual who is the subject of personal data.

De-identified Administrative Data
Administrative 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.

Disclosure
The act of making information or data available to one or more third parties.

Disclosure Control
A technique used to control the risk of individuals being identified from statistical data – typical methods include removing or disguising data relating to individuals with unusual sets of attributes.

Indexer
Individual (or body) who receives personal data from one or more data holders and determines which records in each dataset relate to the same individual (or entity). The indexer then creates a unique reference for each individual (or entity) and a corresponding key to allow the data from the different sources to be joined. The reference is provided to the data holder(s) who then sends the reference and additional data to the linker. The indexer sends the key to the linker to carry out the linkage on behalf of the data holders. Only information which is required for the linking process is passed to the indexer.

Individual Reference / Identifier
A sequence of characters and / or numbers that is used and / or assigned by an organisation to a person to identify uniquely the person for the purposes of the organisation's systems and operations. A Persistent Identifier is an identifier that will remain the same regardless of where the identifier is located, for example, one that is used in several independent databases.

Linker
Individual (or body) who receives data from data holders and links them together using a key created by the indexer.

National Administrative Data
Data derived from information collected and maintained as part of a national administration system, such as health records, vehicle licensing, tax and social security systems.

Open Data
The government's recent white paper defines Open Data as:

Data that meets the following criteria:

- accessible (ideally via the internet) at no more than the cost of reproduction, without limitations based on user identity or intent;
- in a digital, machine readable format for interoperation with

other data; and

- free of restriction on use or redistribution in its licensing conditions.

Personal Data
Data which relate to a living individual who can be identified

(a) from those data, or

(b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller,

and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual. To determine if data are personal data, refer to the flowchart on page 6-7 of the Information Commissioner's Office What is personal data? – A quick reference guide.

Pseudonymised Data
Such data 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. Pseudonymised data can often, but not always, be used in place of identifiable data.

Secure Data Access Facility (Safe Haven, Safe Setting)
Has traditionally meant a physical environment where access to disclosive data can be controlled, e.g. a 'safe room'. With the development of software to provide similar levels of information security, safe settings can also be used to denote a virtual environment where users can use a client system to run scripts on data stored on protected computational facilities.

Sensitive Personal Data
Personal data consisting of information as to

(a) the racial or ethnic origin of the data subject,

(b) his/her political opinions,

(c) his/her religious beliefs or other beliefs of a similar nature,

(d) whether he/she is a member of a trade union (within the meaning of the Trade Union and Labour Relations (Consolidation) Act 1992),

(e) his/her physical or mental health or condition,

(f) his/her sexual life,

(g) the commission or alleged commission by him/her of any offence, or

(h) any proceedings for any offence committed or alleged to have been committed by him/her, the disposal of such proceedings or the sentence of any court in such proceedings.

Statistical Information
Information which is held in the form of numerical data, nominal data (e.g. gender, ethnicity, region), ordinal data (age group, qualification level), interval data (month of birth) or ratio data (height, weight, age in months, length).

Trusted Third Party
An organisation which received personal identifiers supplied by data holders for indexing purposes, and supplies linkage details for de-identified data supplied to ADRCs.

Appendix 4 List of acronyms

ADLS	Administrative Data Liaison Service
ADRC	Administrative Data Research Centre
BIA	Business Impact Assessment
BIS	Department for Business, Innovation and Skills
CIS	Customer Information System
CRCA	Commissioners for Revenue and Customs Act 2005
CMO	Chief Medical Officer
DfE	Department for Education
DWP	Department for Work and Pensions
ESRC	Economic and Social Research Council
HMRC	Her Majesty's Revenue and Customs
LS	Longitudinal Study
MCRDC	Michigan Census Research Data Center
MoJ	Ministry of Justice
MRC	Medical Research Council
NILS	Northern Ireland Longitudinal Study
NISRA	Northern Ireland Statistics and Research Agency
ONS	Office for National Statistics
PIA	Privacy Impact Assessment
RDC	Research Data Centre
SAIL	Secure Anonymised Data Linkage
SHIP	Scottish Health Informatics Programme
SLS	Scottish Longitudinal Study
SRSA	Statistics and Registration Service Act 2007
VML	Virtual Microdata Laboratory

Appendix 5 Questions and answers

In the course of its investigations, and in discussions with interested parties as its recommendations evolved, the Administrative Data Taskforce has received various questions. Those which may be of general interest, and the responses from the Taskforce, are reproduced below.

Q1: What are the risks of not establishing the system proposed here?

Answer: The UK is a world leader in the analysis of large-scale survey datasets, which have been used successfully to address a range of academic and policy relevant questions. However, while longitudinal cohort and panel studies or large-scale cross-sectional surveys provide uniquely valuable information, they have some drawbacks including their small sample sizes and rising attrition rates. In many other countries, particularly in Scandinavia and some parts of Europe such as the Netherlands, advantage has been taken of national datasets based on routinely collected government administrative data. Unfortunately, despite having a wealth of such data in the UK, to date the use of these resources has been patchy at best. Failure to make better use of these resources, in a secure and safe way, would undermine UK academic research, and hinder better policy-making and policy evaluation.

Q2: Why not just have one Administrative Data Research Centre for the UK?

Answer: Progress in facilitating research access to and linkage between administrative datasets has been achieved at different rates in the devolved administrations. Scotland has made major advances in this area, in part due to the different legal environment but also because of the high degree of trust that has developed between researchers and data holders. In addition, some of the datasets that exist in each country vary, requiring different expertise to manage them. Establishing just one ADRC would inhibit the gains that have already been made in different countries of the UK, and restrict the centres of excellence required to support the research community.

Q3: Will departmental data holders cede any decision-making powers on access to and linkage between various administrative datasets to the Administrative Data Research Centres?

Answer: No. Data holders in government departments will retain all the decision-making powers that they currently have. The decision to allow data to be passed to an ADRC will be made by data holders on a case-by-case basis in line with their existing procedures, through the project application process to be co-ordinated by the Governing Board.

Q4: Will researchers be able to see identifying personal information on people and organisations (e.g. name, address, dates of birth, residential postcode, National Insurance or National Health Number)?

Answer: No. All identifying personal information on people and organisations will be removed from data before it is passed to the ADRCs. Neither ADRC support staff nor researchers will be granted access to such personal information.

Q5: Will the ADRCs become data warehouses, gradually accumulating more and more administrative datasets as time goes on?

Answer: No. The ADRCs will facilitate research access to administrative data on a case-by-case basis and will not act as data warehouses, storing increasing numbers of national datasets. However, in some circumstances (and with the explicit advance approval of data holders) research datasets (subsets of national datasets relevant to the specific research question) may be archived by an ADRC to facilitate the replication or extension of research at a later date. Applications to work on such historical datasets would require a full approval process as for a new research project.

Q6: Will it be possible to replicate and / or extend research based on access to and linkage between datasets?

Answer: Researchers who work on administrative datasets within an ADRC will be required to document the data they have processed and to make available for reuse the statistical syntax that they have generated in the course of their research. These metadata will facilitate replication or extension of their research. Libraries of 'useful' quality assessed code for analysing administrative data (e.g. for creating commonly used variables from raw data) will be held in an open ADRC archive and the ADRC support teams will provide expertise on how best to take advantage of these resources.

Q7: How will the security procedures in the ADRCs be monitored?

Answer: Each ADRC will be required to implement strict security procedures based on approved international standards to prevent any unauthorised access to data, inappropriate use of data or the extraction of data other than in summary form. Monitoring of these procedures will be achieved via independent audit conducted in accordance with international standards of data security, which will be commissioned by the Governing Board.

Q8: Can other access and linkage centres sit alongside the proposed ADRCs?

Answer: Yes. Other access and linkage centres will play a vital part in improving research opportunities. The Office for National Statistics Virtual Microdata Laboratory, HMRC's Datalab and the linkage functions provided within the DoH's Health and Social Care

Information Centre are examples where access and linkage across datasets within specific sectors or areas of activity have been established. The proposed national ADRCs, which will manage linkage **between** departments, will be expected to work closely with these sector-based administrative data centres to provide the most efficient access arrangements where agreement for access and linkage has been approved by data holders.

Q9: Will researchers be provided with datasets including individual records by the ADRCs?

Answer: No. The three access methods do not allow raw data to be released from the ADRCs. The researcher in the remote safe setting will be analysing data in the ADRC safe setting remotely, so no raw data needs to pass between these locations. Only the results of statistical analyses (e.g. contingency tables, regression parameters etc.) undertaken using the data held by the ADRCs will be supplied to the researchers.

Q10: Does a researcher have to approach a specific ADRC, or can they choose?

Answer: They can choose. The choice of the ADRC will depend upon the expertise available within an ADRC, the workload involved and the type of access granted to the researcher. It is possible that each ADRC will become proficient in providing access to specific datasets and that this proficiency will guide applications directed to a specific ADRC. For those studies requiring data for more than one country, or for the UK as a whole, each ADRC will be capable of providing support. Once datasets have been produced, the researcher can access the data at the relevant ADRC or from one of the virtual safe settings.

Q11: Who will run the whole system?

Answer: The whole system will operate with the support of data holders and those who are funding the system. The Governing Board will report annually to a body responsible to Parliament. This report will form the basis of effective monitoring of the system and will propose ways in which the system can be further enhanced.

Q12: Will data from the private sector be held in ADRCs and will private sector researchers have access to government data via the ADRCs?

Answer: While there is considerable research value in many of the large datasets held by private sector organisations (e.g. customer databases, service records, financial transactions), these usually have commercial value and may be subject to specific controls on their use for research by the organisations which control such data. The current arrangements proposed in the report of the Taskforce will not allow access and facilities for the private sector. These are important issues which will be addressed by the Governing Body as a matter of priority.

Q13: If generic data linkage legislation is not passed, will anything change?

Answer: Yes. There is much to be gained from what is termed the 'dual track approach' including setting up the proposed ADRCs, providing mechanisms for data holders to make consistent decisions regarding data access, providing resources to enable this purpose to work efficiently and promptly, and to promote the strategic work of the Governing Board. Although the existing problems of access to specific administrative datasets will be eased if generic data linkage legislation is passed, there will be considerable improvements in the scope for and efficiency of research based on linked administrative data without such legislation.

Q14: Once ADRCs are established will all linkages of government data, including those within specific sectors (such as linking GP and hospital data in the health sector), have to be undertaken there?

Answer: No. The aim of the ADRCs is to provide the environments within which data can be linked in the most efficient manner. If, for example, the research in question only requires access to linked health records, this could take place within the National Health Service Information Centre for Health and Social Care or the Information Services Division of NHS National Services Scotland. If the linkage requires cross-departmental co-operation (e.g. linking health records to social security records) the relevant ADRC will collaborate with the sector specific data linkage centre to achieve this link.

Q15: How will data be passed to ADRCs from government departments? Is there a risk of data getting lost during this process?

Answer: No. Data will be de-identified (personal identifiers removed) before subsets of requested data are passed to the ADRCs. These de-identified data records will be encrypted before transfer by methods which minimise the risk of loss. The same procedures apply to the transfer of identifiers to and from third party linkers. Such data are already passed between some departments using secure methods.

Q16: If a researcher is found to be attempting to act unscrupulously what penalties will be enforced and under whose authority?

Answer: Where the data concerned are covered by the Statistics and

Registration Service Act 2007 (SRSA) or the Commissioners for Revenue and Customs Act 2005 (CRCA), abuse of the conditions of access may constitute an offence under these acts, for which the offender is liable on conviction on indictment to imprisonment for a term not exceeding two years, or to a fine, or both. Where the data are not covered by the SRSA or the CRCA, the penalties for abuse of the conditions of access will range from a ban on future access via the ADRCs for the offender for a specified period, a ban on access for his / her institution and possible liability for fines levied by the Information Commissioner's Office.

Q17: How will the proposed system be more efficient than the current system?

Answer: The most inefficient part of the current system relates to researchers being prevented access to linked data from different departments and the delays that researchers experience when applying for access to administrative data. The proposed system will allow for more high-quality research to be achieved using administrative data than is currently the case. Research applications will be made more speedily and data access will be provided in an environment dedicated to research. Best practice will be disseminated and specific access and linkage problems will be identified and addressed with dedicated resources.

Q18: How will the Governing Board for the ADRCs perform both its operational and strategic development functions?

Answer: The Governing Board is set up to ensure the smooth running of the ADRCs. We envisage it meeting up to four times a year initially to review progress and to assist with the resolution of specific problems associated with particular research applications. According to the demands placed upon it by the research community, it will establish structures and evolve practices that allow it to conduct its operational work in the most efficient manner. The strategic work of the Governing Board may require it to commission studies, undertake enquiries and possibly co-opt specialist members for specific periods. A sub-group will be responsible for assessing applications to access data in the ADRCs. Resources will be provided to the Governing Board for these purposes.

Q19: If the Governing Board refuses an application, can the researcher continue to request data directly or through other means?

Answer: There is nothing to prevent a researcher taking such action. However, the grounds for refusal will be made clear, making it unlikely that further application for access via alternative routes would be pursued.

E·S·R·C
ECONOMIC & SOCIAL RESEARCH COUNCIL

The Economic and Social Research Council is the UK's leading research and training agency addressing economic and social concerns.

We aim to provide high quality research on issues of importance to business, the public sector and government.

The issues considered include economic competitiveness, the effectiveness of public services and policy, and our quality of life.

The ESRC is an independent organisation, established by Royal Charter in 1965, and funded mainly by the government.

www.esrc.ac.uk

MRC Medical Research Council

For almost 100 years the Medical Research Council has improved the health of people in the UK and around the world by supporting the highest quality science. The MRC invests in world-class scientists. It has produced 29 Nobel Prize winners and sustains a flourishing environment for internationally recognised research. The MRC focuses on making an impact and provides the financial muscle and scientific expertise behind medical breakthroughs, including one of the first antibiotics penicillin, the structure of DNA and the lethal link between smoking and cancer. Today MRC funded scientists tackle research into the major health challenges of the 21st century.

www.mrc.ac.uk

wellcome trust

We are a global charitable foundation dedicated to achieving extraordinary improvements in health by supporting the brightest minds.

www.wellcome.ac.uk