The UK Administrative Data Research Network – its genesis, progress and plans for the future.

Peter Elias, Institute for Employment Research, University of Warwick

Manuscript submitted 9 May 2017

Abstract

This paper describes the ways in which UK administrative data are becoming more widely available for research. It outlines the historical context of these developments, details the networked infrastructure that has now been put in place, and depicts the continuing legal measures that are required to bring this network to fruition. It focusses on the lessons that have been learnt as work has progressed, drawing on this experience to elucidate some principles that may have relevance to similar attempts to promote better access to and linkage between administrative datasets in different cultural and legal settings.

Key words

Networked data infrastructure; administrative data; data access; data security

Short biography

Peter Elias CBE is a research professor at the Institute for Employment Research, University of Warwick, UK. His research interests include higher education and the labour market, processes of occupational change, and the evaluation of large scale labour market programmes. Over the past ten years he has worked closely with the UK Economic and Social Research Council, being actively involved in the development of research data infrastructure within the UK.

1. Introduction

This paper reports on research infrastructure developments in the UK, designed to facilitate and improve record level access to administrative data. It describes activities over the last two decades as access for research purposes has moved from a situation where it was unreliable, haphazard and unregulated, to one where all those involved, from data holders to the general public have a clearer understanding of the needs of the research community and through which researchers can plan their research and be confident in the knowledge that access to administrative data will be forthcoming if the conditions upon which access is granted are fulfilled.

The aim of this paper is not to portray these developments as a specific model for others to emulate. Rather, it seeks to highlight the problems that have been addressed in establishing this new access regime and to derive the lessons learned. What were the main obstacles to improve access and how were they tackled? Does the new system provide researchers with access without placing a large burden on them? How does it sit alongside other structures that have been set up to provide research access to specific datasets? What has it cost so far and how does this compare with the scientific benefits expected from this infrastructure?

The structure of the paper is as follows. Following this introduction, Section 2 describes earlier attempts to improve person level access to administrative data in the UK. What was accomplished, how successful were these attempts and how did this translate into the approach adopted in 2012 when work began in earnest to establish the Administrative Data Research Network? Section 3 describes the task force that was set up to examine the barriers to access that researchers had faced when seeking access, recommending how

these barriers could be lifted through the creation of a new networked infrastructure. Section 4 details the way in which these recommendations were enacted through the creation of the Administrative Data Research Network (ADRN), looking at what it has achieved in its first two years of operation. This section draws also on recent and independent reviews of its activities. What are the main problems it now faces and what should be done to improve the Administrative Data Research Network, thereby further realising the research value of administrative data?

Section 5 examines in some detail how a better understanding of the needs of data providers has impacted upon the early work of the Administrative Data Research Network. Section 6 highlights the lessons learned. It draws on this experience across the UK to signpost the ways in which others might chart a roadmap towards improved access to administrative data. But the barriers to access and data linkage that researchers are likely to face will be legal and may also be 'cultural', in the sense that data holders have over many years developed particular approaches to data sharing for research. Such hurdles are both country and agency specific, as will be made evident from the differing degrees of success in developing data access arrangements across the four countries of the UK and the slow rate of progress towards access to data held by particular government departments. Bearing this in mind, the lessons taken from the UK experience can only be drawn at the level of principles, not specifics. These principles are enunciated in the concluding section.

2. Earlier attempts to improve access to administrative data

The digitisation of government administrative records commenced in the UK in the 1960s with the advent of mainframe computers, card readers and line printers. Access to the data held in these computers was designed for administrative purposes only. Information

derived from such data and placed into the public domain was aggregated, usually across time periods, and was generally months out of date at the time of publication in official statistical gazettes. Access for research purposes unrelated to these administrative functions was generally not regarded as a legitimate function for those collecting and processing such records.

In the late 1970s and through the following two decades, government departments and agencies began to appreciate the research value inherent within their large and growing databases and started carrying out research, usually to inform service delivery, using unit level statistical records for which they had responsibility. These activities were either conducted 'in-house' by their own research teams or were 'bought in' services provided by contracted researchers, often involving university research teams given their expertise in computer programming and statistical analysis. Access for external research purposes was rarely granted, a situation which was further confounded by the fact that there is no centralised statistical authority within the UK that could provide a more regulated approach over access to person level administrative data. Most departments of government retain such control, except for information deemed to be 'national statistics' (a term that refers to an aggregate statistic which achieves a certain level of quality, defined by a Code of Practice, rather than raw data *per se*). Some departments of government cover the whole of the UK (*e.g.* taxation, social security, defence), others are specific to each country of the UK (*e.g.* education, justice, health). To complicate matters further, there are three national statistical authorities, each covering population, the economy and society (see Box 1).

Box 1: The National Statistical System in the UK

The United Kingdom has devolved responsibilities for the collection, processing and publication of statistical information to its constituent countries.

In Scotland, the national statistical authority is *National Records Scotland* (https://www.nrscotland.gov.uk/statistics-and-data/statistics).

For Northern Ireland, these tasks are the responsibility of the *Northern Ireland Statistics and Research Agency* (http://www.nisra.gov.uk).

In England and Wales the *Office for National Statistics* (https://www.ons.gov.uk) has responsibility for both countries, as well as acting as the coordinating agency for UK-wide statistical information.

By the turn of the century, many researchers were becoming aware of the research potential of the vast amounts of data that had been accumulating within government departments. For their part, government departments and agencies were being urged to engage in cross departmental data sharing to improve service efficiency. On the research front, reports commissioned as part of the UK National Data Strategy[1] detailed numerous datasets that could prove to be valuable research assets if access procedures could be improved (Smith *et al.,* 2004; Jones and Elias, 2006). These reports listed of some of the more successful efforts that had been made by researchers, both within government and academia, to improve research access to particular types of data (See Box 2 for an example).

Box 2: The National Pupil Database

The National Pupil Database (NPD) is described as 'one of the richest education datasets in the world, holding a wide range of information about students who attend schools and colleges in England' (See http://www.adls.ac.uk/department-for-education/dcsf-npd/?detail). The NPD combines annual examination results of pupils with information on pupil and school characteristics and is an amalgamation of a number of different datasets, including attainment data from key stages in a child's education, and Schools Census data which are linked using a unique identifier for each pupil.

These data are held by the Department of Education in England. Access was promoted through a user group that was set up and previously (2006-2013) organised by Simon Burgess, Director of the Centre for Market and Public Organisation (CMPO) and Harvey Goldstein, Professor of Social Statistics at the Centre for Multilevel Modelling.

Despite the difficulties that some have experienced, the user group has been successful in promoting research access to these data. More than 80 publications were listed by the Administrative Data Liaison Service (http://www.adls.ac.uk/ADLS-hub/publications/?dept=DfE&ds=229&search_submit=1)

An influential report published in 2008 stressed the need for data holders to have a better understanding of their legal powers to share data, given that this was widely recognised as a major obstacle (Thomas and Walport, 2008).

Encouraged by these reports and the growing demand for access to administrative data for research from the academic research community, in 2006 the UK Economic and Social Research Council (ESRC) agreed to fund both an Administrative Data Liaison Service (ADLS) and a Secure Data Service (SDS). These were established in 2008. The former did not provide access to administrative data, instead it worked as a repository for metadata relating to the wide variety of datasets held be government departments and of potential interest to social scientists. The Administrative Data Liaison Service rapidly became a 'first stop shop' for researchers seeking to discover whether or not any administrative data were potentially available that could meet their research needs. The SDS was set up to mirror a similar secure data laboratory established by the Office for National Statistics[2], initially

holding unit level records for organisations on the national business register. It has since expanded its holdings and developed remote access technology to the highest security levels required by government departments.

The success of these services was notable. Awareness was raised within the research community of the research potential of the wide variety of administrative data held by government departments. Secure access to such data could now be guaranteed. Despite these efforts, progress on data sharing and record matching across different data sources and in different parts of the United Kingdom was patchy. In part this was a consequence of legal barriers – some very real and others perceived by data guardians in their attempts to interpret complex laws – and of the unsatisfactory procedures which had evolved in an *ad hoc* manner to respond to requests from researchers for access to unit level data. In Scotland, with a different legal system and where there existed a high degree of trust between data guardians, significant progress in linking individual level data was being made[3]. This was not the case in England and Wales. While there was a general understanding among data holders and researchers of the need to protect data subjects from any misuse or misappropriation of data about them, mistakes had sometimes been made due mainly to errors in data processing or through the use of sub-contractors providing data processing services. For example, for an evaluation of a particular employment programme, individual level data on many thousands of programme participants were supplied under contract to a university based research team, complete with names and addresses of all participants.[4] These errors, together with a number of widely reported losses of millions of personal data records[5], led to a situation where all government departments took steps to minimise any exchanges of personal data with other agencies. Research use of administrative data came to a virtual standstill in many

departments from 2008.[6] Nonetheless, the Administrative Data Liaison Service had been established and the Secure Data Service[7] was open for business at this time. The stage was set for change.

2. The Administrative Data Taskforce

In September 2010, a new chief executive took up post at the Economic and Social Research Council. With his background in longitudinal data linking, based in the more progressive data access environment that existed in Scotland, one of his first acts was to revitalise efforts to improve access to administrative data across the UK. The first major breakthrough came In May 2011. At a meeting convened by the Wellcome Trust, senior social scientists, research funders, data holders and government statisticians met to discuss the scientific value of linked administrative records[8]. It quickly become clear that, in the medical sciences, new legislation coupled with the centralisation of digital records from various parts of the National Health Service, was revolutionising the opportunities for ground breaking research. Strategic plans were developed for the exploitation of such data. This was not the case for a range of administrative data of interest to social scientists (*e.g.* tax, social security records). Consequently, a task force was established to determine why this was the case and to make recommendations to improve research access to such data. Membership of the task force was drawn from data guardians within government departments, research funding organisations, the private sector and the Information Commissioner's Office[9]. The Terms of Reference for the task force are shown in Box 3. The task force worked to address its terms of reference by focussing on three areas: legal and ethical challenges to data access; public acceptability of improved access for research; and models for data access and linkage. It was also tasked to advise on the cost implications of recommendations for improved access.

The Administrative Data Taskforce (ADT) worked to a rapid timetable. Within one year of its formation in 2011 the task force reported its findings (ESRC, 2012)[10], noting in particular that even in those instances where successful data linking and sharing of administrative data had been achieved, the process was often very slow[11]. There was evidence that a number of funded research projects had been abandoned due to the lack of support from data holders after access had been agreed in principle.

The report made a number of important recommendations relating not just to the provision of new research infrastructure but, more importantly, issues about the governance of research access, the legal situation, the ethics of data sharing for research and the need to build trust between researchers, data holders and the general public. Specifically, this translated into:

- the establishment of new facilities in each of the four countries of the UK (to be called Administrative Data Research Centres (ADRCs)), together with a UK governing board and a UK information service (to be called the Administrative Data Service (ADS)) as the point of entry for research applications to gain access to administrative data. These research centres and the new service were to form the backbone of a new Administrative Data Research Network;
- a call for new legislation to clarify the legality of data sharing between departments and with others particularly for those agencies established by statute and without expressed powers for data sharing;
- the creation and implementation of a single UK wide accreditation process for researchers wishing to gain access to administration data;
- a strategy for public engagement;

- sufficient resources to provide the above.

Box 3: Administrative Data Taskforce (ADT)

The ADT seeks to propose new mechanisms and collaborative agreements to enable and promote the wider use of administrative data for research and policy purposes.

Terms of reference

The key role of the Taskforce was:

- identification of potential benefits and risks from increased research use of administrative data;
- the development and introduction of common procedures to provide more efficient research access to administrative datasets;
- clarification of the legal situation governing the research 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 research access to administrative data for research;
- the development of agreed methods for data linkage;
- procedures to raise public awareness of the benefits arising from research use of administrative data and data linkage;
- identification of likely resource implications arising from increased research use of administrative data.

Some secondary priorities for the Taskforce could include:

- further development of a 'metadata authority' to assemble and disseminate information relating to the research 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;
- guidance on data access, including the use of 'safe-settings' and how the research environment should be controlled;
- proposals for how the quality of administrative data may be assessed.

The government response was supportive (Department for Business, Innovation & Skills, 2013)[12], essentially endorsing the recommendations of Taskforce. Funding to facilitate implementation of the recommendations was announced. Shortly after funding was announced from Department for Business, Innovation and Skills, a technical group was formed to advise the ESRC on the appropriate technical and infrastructure design and published a report (Elias, 2013)[13], leading to the commissioning of four administrative data research centres, one in each country of the UK, together with an advisory, liaison and

coordination service (encapsulating the Administrative Data Liaison Service set up five years previously and called the Administrative Data Service).

The usual approach to commissioning research infrastructure is for the process to be managed by the research councils on behalf of the science ministry releasing the funds. In this manner independence from government funding is maintained while the scientific merit of various bids to construct the infrastructure is evaluated by senior scientists who are vetted to ensure they have no affiliation to those bidding. A different model was required in this instance. The cooperation of government departments was vital to the success of the Administrative Data Research Network, and there was a role to be played also by the various statistical authorities of the UK. To this end, the three statistical authorities of the UK were invited to state how they would work with the centres once these had been commissioned and to produce a bid for funds for such participation. The major government departments were invited to nominate representatives to join the commissioning panel, which was chaired by an academic and included international expertise.

Commissioning of the four centres and the service was completed by the autumn of 2013. The big question now was 'would it work'?

3. The Administrative Data Research Network

The decision to commission one administrative data research centre in each country of the UK was more than simply a gesture towards the increasingly devolved nature of the four countries of the UK. It recognised that progress in developing research access had been moving forward at a better pace in Scotland, Wales and Northern Ireland than was the case in England[14]. This was not a consequence of the different legal systems in these countries, but more a reflection of their smaller and more integrated administrative departments of

government. The groundwork performed for the Taskforce had established that knowing the people involved in data sharing activities and having trust in them was a major factor in facilitating data access and linking. One of the aims of the newly established Administrative Data Service was to explore these differences and to determine how best practice could be spread across the network.

As the new centres and the coordinating Administrative Data Service were being set up, agreement was reached with the UK Statistical Authority (UKSA – an independent body operating at arm's length from government, accountable to UK parliament and with oversight responsibilities for the three statistical agencies of the UK) to provide a home for and to service the Governing Board for the network. Given its independence from government this gave the network a degree of authority as an official independent UK body without it having to be established via an Act of Parliament.

The Governing Board met first in April 2014 and immediately identified a number of strategic issues for the board members to consider, including[15]:

- Public confidence;
- Legal issues;
- Demand and capacity;
- Data quality.

On the legal front, progress was much slower given the complexity of the legal issues to be addressed and the need to ensure that the privacy lobby in the UK (an amalgamation of interest groups with concerns about the potential invasion of privacy that data sharing and matching could foster) was consulted and that their concerns could be reflected in the proposals being developed. The Cabinet Office (the department responsible for the

development of new legislation) began a process termed 'Open Policy Making' in April 2014, concluding this in April 2015 with the publication of a policy paper (Cabinet Office, 2015)[16]. This paper recommended that 'any public body that needs it (should have) the necessary power to engage in trusted third party data shares with one or more sources for the purpose of research' and '(this power) should only be used when all the bodies involved in the data share, other than the data sources, are accredited bodies'. This is a vital power for those administrative departments of government and agencies established by statute, given that they have no power to share data for research unrelated to their administrative functions.

In the spring of 2016 the Cabinet Office published a consultation on the potential section of a new Data Sharing Bill and included clauses informed by the Open Policy Making discussion to improve access to administrative data for research purposes. The consultation responses provided the necessary confidence for Ministers to prepare a Bill to be presented to Parliament. The Bill was placed before Parliament in July 2016. This Bill, known as the Digital Economy Bill, is broader and includes many aspects of the digital activity of government and its arms lengths bodies, but sections designed to facilitate sharing for research purposes and also sharing for statistical purposes have been included. These sections have been subject to scrutiny as the Bill progresses through the UK parliament. It has now passed through the committee stage in the House of Commons and is in progress through the upper chamber. Royal assent is anticipated for mid 2017.

4. Reviewing the work of the Administrative Data Research Network

The Administrative Data Research Network has made considerable progress by bringing a coherent approach towards access to and the use of administrative data for research. The

composition of the Governing Board means that issues identified can be acted upon quickly and communicated back to data holders if necessary. Concerns about data security can be addressed in a manner which is acceptable to all data providers. Training is given to researchers who wish to gain access to secure data (see Box 4). The Board has formed an Approvals Panel, consisting of data holders, senior academics, lay members and an expert in data privacy and confidentiality, which acts to vet all research proposals planning to use the Administrative Data Research Network as a resource. The Approval Panel undertakes its work in accordance with the following criteria:

i. an outline of an acceptable process for review of the ethical implications;
ii. clear understanding of and approaches to address any potential privacy issues;
iii. a feasible project for being undertaken in the Administrative Data Research Network with the data proposed;
iv. a project of sufficient scientific merit to warrant the use of Network resources;
v. a convincing case is made for potential public benefit.

The fifth of these criteria reflects the fact that, for most of the applications to undertake research using administrative data, the data subject may not have consented to such use.

The Board also informed the establishment of what is named the National Statistician's Data Ethics Advisory Committee[17] to ensure that ethical approval can be given for research proposals from Government departments and third sector organisations. Unlike universities, most bodies such as these do not have any ethics committees set up to review and provide ethical safeguards for the subjects of research.

The Administrative Data Service, with the four Administrative Data Research Centres, has been receiving applications for research, approving proposals, and training researchers in procedures for access to data in a secure setting. The network has promoted the work of the network through YouTube videos, special events held around the UK, establishing an annual conference, as well as critically ensuring that the network data security conforms to the standards required by data holders. By December 2016 the network had received over 50 project requests, with three being rejected by the Approvals Panel[18]. Only two had been completed by this date.

Having been in operation for two years, the process for reviewing the work of the Administrative Data Research Network commenced in mid 2016. Although the network had been under close scrutiny by its governing board, all major infrastructure projects funding from UK capital funding are subject to a formal and independent review process known as 'Gateway Review'[19]. Earlier reviews had focussed upon the progress made by the Administrative Data Research Network in establishing the infrastructure. The review in mid 2016 centred upon the delivery of the benefits realised by the investment. It concluded that 'key risk relates to the availability of UK wide data'[20].

Box 4: National Training for Accreditation of Researchers

Before they can access sensitive and/or potentially disclosive data via the Network – or indeed any other major UK data services – researchers have to undertake the National Training for Accreditation of Researchers, including passing a post-training test.

The training has been developed by a consortium of UK data services:

- Administrative Data Research Network
- HMRC Data lab
- ONS Virtual Microdata Laboratory
- UK Data Service (Secure Lab).

Other data services are expected to join the programme in the future.

Trust in researchers is essential, so the programme has been designed to make sure they behave safely and responsibly when accessing detailed data. With this in mind the training covers two main areas:

- The legal responsibilities (and expected behaviour) of researchers accessing detailed data. They learn about the 'five safes' – safe researchers, safe projects, safe settings, safe outputs, and safe data – as well as the legal and procedural background to UK data access and their role and responsibilities in data protection.
- Output disclosure control. In this module researchers learn about Statistical Disclosure Control, how a disclosure might happen and how to avoid disclosive outputs.

Source: https://adrn.ac.uk/media/1190/newsletter-final_web.pdf

This risk derives from the difficulties that researchers have continued to experience in gaining access to linked data from a number of key government departments holding UK wide data on individuals, including tax records, social security records, driver and vehicle licence records. Within the devolved parts of the UK (Scotland, Northern Ireland, Wales) the network was performing well. But without similar success with these important UK wide sources of administrative data, it would not achieve the vision set out in the Taskforce report in 2012. What was the problem? The answer to this question requires a deeper understanding of the needs of data providers.

5. Understanding the needs of data providers

Faced with a request to allow research access to data, data providers will consider four factors: the legality of the request; the public benefit (including the ethics of using the data they hold for a purpose which does not align with any consent for reuse given by the data subject); the resources involved in meeting the request; and the risks that they are exposed to if they meet the request. Each of these factors is examined in turn.

Box 5: Data Protection Legislation in the UK

In the UK, the principal legislative instrument governing the collection, sharing, transmission and processing of electronic records is the Data Protection Act 1998. This set out 8 data protection principles:

1. Personal data shall be processed fairly and lawfully.
2. Personal data shall be obtained only for one or more specified and lawful purposes, and shall not be further processed in any manner incompatible with that purpose or those purposes.
3. Personal data shall be adequate, relevant and not excessive in relation to the purpose or purposes for which they are processed.
4. Personal data shall be accurate and, where necessary, kept up to date.
5. Personal data processed for any purpose or purposes shall not be kept for longer than is necessary for that purpose or those purposes.
6. Personal data shall be processed in accordance with the rights of data subjects under this Act.
7. Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data.
8. Personal data shall not be transferred to a country or territory outside the European Economic Area unless that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects in relation to the processing of personal data.

The second principle would appear to preclude reuse of administrative data for research purposes, but the Act also states that re-use for research is permissible provided:

- that the data are not processed to support measures or decisions with respect to particular individuals, and
- that the data are not processed in such a way that substantial damage or substantial distress is, or is likely to be, caused to any data subject.

Such data can be processed only for research purposes and can be kept indefinitely if:

- they are processed in compliance with the relevant conditions [as above], and
- the results of the research or any resulting statistics are not made available in a form which identifies data subjects.

Many would regard the legality of a request as a key obstacle to data sharing. Data protection laws are designed to provide data subjects with safeguards against any inappropriate use of their data. In so doing the concept of 'personal data' is invoked, where personal data are defined as any data relating to a living person which identifies them or has the potential to identify them given the detail they contain. Data can be legally shared for research purposes if the risk of revealing the identity of an individual is minimised. The legal basis for sharing data then hinges upon how this risk is assessed and whether or not the risk of disclosure of identities has been minimised (see box 5).

There are some data types for which the legal basis for sharing is much more explicit than is established via data protection legislation. In the UK, some data providers are organisations established by statute. In other words, they exist because of a body of law which states what they can do. Anything not listed as a lawful activity of a statutory body is *ultra vires.* For example, the UK taxation authority, Her Majesty's Revenue and Customs (HMRC), was established by an act of parliament in 2005[21]. This act allows Her Majesty's Revenue and Customs to disclose information to other organisations or persons when the disclosure is 'made for a purpose of the Revenue and Customs' (Commissioners for Revenue and Customs Act 2005, section 18, subsections (1) and (2)). Given this stipulation, Her Majesty's Revenue and Customs can only release data for research if the research can be construed as a purpose of Her Majesty's Revenue and Customs, *e.g*. improving the efficiency of tax collection methods. This is a severe limitation that has had to be addressed via a change in the law. For non-statutory bodies, such as the Department for Work and Pensions, they already possess the right to share data[22], but have shown a degree of reluctance in this respect.

The issue of consent from data subjects for re use of social and economic data collected for a specific administrative purpose has not figured as a restricting factor by data holders. It is generally recognised that, as long as the research that forms the basis of the data request has some public benefit, then it is in the wider public interest to reuse data for this purpose[23]. The problem is how to define 'public benefit'. This can only be assessed on a case by case basis, to examine the potential benefits that could derive from research findings, weighing these against the privacy rights of the individual or organisation whose data are to be reused for proposed research. The Administrative Data Research Network effectively hands the task of conducting this balancing of benefits and rights to its Approval Panel, who in turn request an initial assessment from the research proposer. This situation may change with the adoption by all member states of the European Union of a new regulation on data privacy[24] which came into effect in May 2018. Recognising the need to ensure that the work of the Administrative Data Research Network is trusted by the public, much work has been done to promote the work of the network and its value whilst ensuring that data about individuals and organisations are kept secure and handled in ways which ensure confidentiality is maintained.

The resources involved in meeting a request for data access are not trivial. The data holder must have knowledge about the conditions under which data will be held and analysed, and must be confident that the data analysts will proceed with due diligence, paying particular attention to the scope for inadvertent disclosure of identities. Some sub setting may be required for data extraction and procedures agreed for safe data transfer, both of which are may require significant resources. The role of the Administrative Data Research Network is to streamline this process and ensure efficiencies which minimise these costs as far as the data provider is concerned by providing secure data transfer, ensuring that data are held in

safe centres, that projects and researchers have been vetted and approved, and in some cases providing resources to data holders to assist with data extraction, documentation and data transfer.

The major difficulty facing any public sector data provider is the evaluation of the risks to which they will be exposed if they grant access for research. Data holders in the UK have a statutory responsibility to prevent the disclosure of the identity of individuals for whom they hold data, except in certain defined situations (see box 5). While the Administrative Data Research Centres and the Administrative Data Service operate procedures to minimise such risk through the use of safe data centres, remote secure access technologies and the use of trusted third parties for data linking and output vetting procedures, data holders may still consider the risk too high. One small error can be perceived of as having powerful negative consequences for the public trust in that organisation. Understanding and meeting this need has been one of the main functions of the Administrative Data Service.

Given the efforts that have been made by the Administrative Data Research Centres and the Administrative Data Service to minimise the risk of disclosure it appears illogical for data providers to have concerns about disclosure risks. However, here is another risk that a data provider may perceive when agreeing to make data available for a research project over which it has no direct control and which has arisen from an independent source. Although an independent Approvals Panel may have given the green light for a specific project, and ethical clearance has been obtained, the data holder may take the view that a particular piece of research may reflect negatively upon its own performance. Simply refusing to provide data for a specific research project could draw unwarranted attention to the data

provider. If this is the case, the best perceived course of action for the data provider is to restrict or withhold access to all data to the Administrative Data Research Network.

The Board of the Administrative Data Research Network has reacted vigorously to the report arising from its Gateway review. Specific government departments have been identified as data owners who may require additional assistance, requesting timetables for delivery of data to projects already approved and providing extra support to these departments if deemed necessary.

Taking stock – what has been created and what has been learnt?

For any research infrastructure, taking stock just two years after its inauguration can only be partial. The infrastructure is maturing and could take up to five years to reach its full potential. With new legislation in the pipeline which may affect its operation, and plans in place to address data access issues, this assessment of what has been achieved must be regarded as provisional.

The first question to ask is the counterfactual – what would the research community now have if this new facility for access to administrative data had not been created? The answer to this is clear. For some data types, there was already an established network, user group or set of procedures that could provide research access. An example of this was given in Box 2. For these data sources, researchers may argue that they now face a more bureaucratic access regime. But many such data sources are accessed via a closed network, not open to researchers outside of favoured institutions. To gain access to the data those favoured researchers often had to sign a contract with the data owning department and sit within that department so there were limitations and barriers even to them.

The role of the Administrative Data Research Network is to improve on the methods and procedures for access by providing added value, particularly through the possibility of linking to other data sets held by different departments and agencies, the improvement of metadata, the provision of training and the provision of the supportive research environment the Administrative Data Research Network can provide. New safe access facilities and investment in technical infrastructure would not have happened without the Administrative Data Research Network. The network is sharing knowledge and expertise and, crucially, providing a repository for this information as administrative data change with the changes to the administrative processes that generate them. So, the answer to the first question is that the research community has an improved environment for research and analysis of administrative data.

Is a cost/benefit analysis of the infrastructure development feasible? This is a more difficult question to address. The costs are known and lie in the region of $20 million for a five year investment[25]. Institutional support is also provided by the higher education institutions that house the Administrative Data Research Centres and the Administrative Data Service, which may total an additional $5 million in capital spend and staff support. The direct employment generated by this investment is less clear, given that many of the staff who are associated with a research centre of the Administrative Data Service. Each centre lists between 30 and 50 staff, but some are from the contributions of established academic staff who would otherwise be engaged in academic related activities. A best guess would put the direct employment associated with this investment at 30 – 40 full time posts. While this may be an interesting statistic to some, the benefits should not be measured in this way. The benefits are the gain to society in having more effective and efficient policy making through the improved research evidence that the infrastructure is providing. It is as yet too

early to address this issue, but it must remain firmly on the agenda as the work of the network progresses.

Could there have been another way of developing improved access? This is an intriguing question and one which the Taskforce grappled with as it formulated its recommendations to create the Administrative Data Research Network. The most obvious alternative approach was to let the four countries of the UK set up their own arrangements for data access independently of each other. This view was voiced within the Taskforce given that Scotland had already made substantial progress in setting up procedures for access to and linkage between datasets held by Scottish departments of government. However, this approach was roundly rejected given that it would not necessarily promote best practice procedures across the four countries of the UK. It also did little to address the issues relating to access to administrative data which were not the responsibility of the devolved administrations (for example: defence, taxation, social security, UK security and driver/vehicle licensing).

There is a related question about the effectiveness of the design of the coordination and management structures within the Administrative Data Research Network. To establish a 'network' and to allocate the funding provided by the UK government science ministry within a challengingly short timescale (the ESRC would usually take nine months or more to commission research infrastructure of this size but government budgetary requirements constrained this period to four months) a collaborative network of equals was established, with no one director of the Administrative Data Research Network (there are five directors) having seniority over others. Furthermore, no consortium or other agreement was put in place to bind together how the Administrative Data Research Centres and Administrative

Data Service operate. With hindsight, although the speed of commissioning and unique nature of this structure led to these decisions, a network of this size and complexity seeking to change the relationship between data custodians in government and researchers needed strong leadership from a single overall director and agreements designed to bind all parties to such a management structure. The Administrative Data Research Network is succeeding despite this, with steps now being taken to address this weakness by appointing one Administrative Data Research Centres director as the overall director of the network. There have also been challenges in engaging at the most senior level with government departments and negotiating access agreements. The post of a Data Access Negotiator has been created and filled by an individual who has worked with senior civil servants and has in fact come from a data owning department, thus providing another step to address some of the ongoing challenges. If these two posts had been in place from 2013, there would have been challenges but possibly not as many as Administrative Data Research Network has faced. Arguably the lack of a senior leader has held back recognition of the Administrative Data Research Network as a single unified and expert network within government circles.

Has it been successful in achieving its objectives? The answer to this question must be a qualified yes. There are access issues that remain to be resolved, and new legislation to contend with in terms of its implications for research access to administrative data, but the direction of progress is positive. Compared with other major investments in research data, the capital costs are relatively low and the potential for the production of well-informed policies based on the reuse of existing data is high. However, with the limited number of projects currently being facilitated but the Administrative Data Research Network, the ongoing running costs may not be regarded as value for money.

What have we learnt from this experience? Here the focus must be on the most critical elements that have driven forward the successful implementation of the research infrastructure. Four factors stand out; governance, coordination, engagement, and review.

The governance of the Administrative Data Research Network has been a major force in ensuring that it will meet its objectives. The composition of the Governing Board, which is comprised of senior representatives from data holders, funders, key executives from the network, representatives from the statistical authorities of the UK, lay members and senior academics gives the Board a powerful voice. But it is not just the constitution of the Board that has helped. Positioning the Board within the independent UK Statistical Authority, a body that reports directly to the UK Parliament and not to the science ministry means that the Governing Board can escalate problems to the UK Parliament if it finds it necessary to do so. It should be noted also that the board operates without a statutory framework. It has not been set up through a legislative process, relying instead on the power of collective voice to carry out its mission. The Governing Board continues to monitor the progress that the network is making towards implementation of the recommendations made by the Taskforce. Two important issues remain to be resolved, access to the network's resources by the private sector and the issue of data retention. The latter reflects the fact that the Taskforce had concerns that retention and reuse of linked data could lead to a situation where the network was holding vast amounts of personal information from individuals. But most government departments view the continued creation and destruction of linked datasets as wasteful, with no ongoing investment in data assets. With the agreement of data owners this approach to retention is now being explored.

Coordination of the network has been an important aspect of its success. The four centres have a degree of independence, but they are bound by a commitment to share expertise, to adopt a common outward identity on the web, in publications and to adopt the same procedures they use for project approval, output vetting, etc. This would have been difficult to establish without the role of the Administrative Data Service, the service that provides cohesion to the network. The Administrative Data Service represents the four research centres on the Governing Board, thus placing it in a strong position relative to the centres.

Engagement, with the general public, data holders, academics and with law makers, has been exemplary. This is an unusual activity for academics, more used to the production of publications rather than YouTube videos and twitter feeds, but it has proven to be popular, particularly the short animated YouTube videos[26].

Independent review of the work of the Administrative Data Research Network has been most helpful in alerting the Board to the priorities it must make in its governance arrangements, particularly the need for stronger overall direction of the network. The review procedures also identified the need for the Board to re-examine its priorities. Rather than an approach of reacting to what researchers want, there needed to be a stronger focus on key departments and datasets, with development of a strategy for the proactive promotion of certain datasets.

6. Conclusion

The four factors discussed above provide some general principles that transcend the particular situation existing within the UK.

First, and most importantly, the UK experience has indicated the need for any network of this scale and complexity to have a *single dedicated full time director*. This role is critical to

success and requires the competence of an individual with the requisite knowledge and expertise that will command the respect of all members of the network as well as those organisations it must negotiate with. Without such a director, the UK network has had to rely on its Governing Board as the agency to provide such leadership. This should not be the role of a Governing Board, which should instead be keeping its eye on the long term strategic aims of the network, its overall costs and potential benefits.

A second principle relates to the *coordination of the network*. With many different functions, such as training, data security, provision of data and metadata, promotion of the networks activities and active liaison with researchers, there is potential for these activities to vary across a large network. Close coordination helps to provide all users of the network, whether as researchers or data providers, with a common experience.

Related to the need for coordination is the *promotion of the identity of the network*. With no market forces to determine success or failure, it is incumbent of those in receipt of public funds or grant income to ensure that the services they are seeking to provide are recognised as having value by researchers and by the public. Promotion of the facilities that such a network can provide does not come naturally to many academics, but the UK experience has shown that this helps to convince both data holders, researchers and the general public that there is much potential for valuable research through the exploitation of administrative data.

Finally, the role of a Governing Board needs careful consideration. While the Governing Board of the UK Administrative Data network has had to tackle problems that should have been the responsibility of a director for the network, it now is operating as it should to

monitor the long term strategic development of the resource and to ensure that it will provide value for money as a major data research infrastructure.

How these principles can be applied as similar networks are created in different environments will depend upon many constraints, such as those operating through legal systems, via the requirements of funding organisations and via the different types of support that academic institutions can offer. Above all though, strong leadership for a network sits as a priority.

Acknowledgements

I would like to thank Vanessa Cuthill and David Hand for comments and advice on earlier drafts of this paper. Additionally, reviewer advice and feedback was most helpful in improving earlier drafts. The views expressed are the sole responsibility of the author.

References

Cabinet Office (2015). 'Conclusions of civil society and public sector policy discussions on data use in government'. Report on the Open Policy-making Process. London: Cabinet Office.

Department for Business, Innovation and Skills (2013). 'Improving access for research and policy: the government response to the report of the administrative data taskforce'. London: Department for Business, Innovation and Skills.

Hasluck, C., P. Elias and A. Green (2003. 'The Wider Labour Market Impact of Employment Zones'. Report to the Department for Work and Pensions. Coventry: Institute for Employment Research, University of Warwick.

Elias, P. (2013). 'Administrative data taskforce: report of the technical group'. Swindon: UK Economic and Social Research Council.

Economic and Social Research Council (2012). 'Improving access for research and policy: report from the administrative data taskforce'. Swindon: UK Economic and Social Research Council.

Jones, P. and P. Elias (2006). 'Administrative Data as a Research Resource: a selected audit'. Report to the UK Economic and Social Research Council. Coventry: University of Warwick.

Hasluck, C., P. Elias and A. Green (2003). 'The wider labour market impact of employment zones'. Report to the Department for Work and Pensions. London: Department for Work and Pensions.

Smith, G., M. Noble, C. Antilla, L. Gill, A. Zaidi, G. Wright, C. Dibben and H. Barnes (2004). 'The value of linked administrative data for longitudinal analysis'. Swindon: UK Economic and Social Research Council.

Thomas, R. and M. Walport (2008) 'Data Sharing Review Report'. Report to the Prime Minister and Secretary of State for Justice.

Endnotes

[1] In 2004 the Economic and Social Research Council (ESRC, one of seven research councils in the UK and the main funding agency for research in the social sciences) decided to adopt a more strategic approach to the development of data resources for research in the social sciences. The first *National Data Strategy* was published in 2006. This strategic approach has continued, with publication of the third strategy in 2013 covering the period 2013 – 2018 (http://www.esrc.ac.uk/about-us/policies-and-standards/national-data-strategy/).

[2] Known as the ONS Virtual Microdata Laboratory (see https://www.ons.gov.uk/aboutus/whatwedo/paidservices/virtualmicrodatalaboratoryvml)

[3] A good example is the Scottish Longitudinal Study, linking population census and health statistics (http://sls.lscs.ac.uk).

[4] Data supplied for analysis of a programme known as Employment Zones (see Elias, P. (2003) 'Do Employment Zones reduce Unemployment? An analysis based on administrative microdata'. In Hasluck, C., P. Elias and A. Green (2003), *The Wider Labour Market Impact of Employment Zones.* Research Report. London: Department for Work and Pensions).

[5] While these losses were not related to requests for access to administrative data by researchers, they caused such a degree of public consternation that all forms of data movement became severely restricted. See https://en.wikipedia.org/wiki/Loss_of_United_Kingdom_child_benefit_data_(2007) .

[6] Smith, G., M. Noble, C. Antilla, L. Gill, A. Zaidi, G. Wright, C. Dibben and H. Barnes (2004) *The Value of Linked Administrative Data for Longitudinal Analysis.*

[7] Now known as the UK Data Service Secure Lab (https://www.ukdataservice.ac.uk/use-data/secure-lab).

[8] 'Linking Data: new scientific possibilities for the Biomedical and Social Sciences in the Digital Age' Frontiers meeting, Wellcome Trust, London, 16-17 May 2011.

[9] The government agency responsible for the enforcement of the UK Data Protection Act 1998, and also responsible for Freedom of Information.

[10] Economic and Social Research Council (2012) *Improving Access for Research and Policy: Report from the Administrative Data Taskforce.* Swindon: Economic and Social Research Council.

[11] An example of this is the data sharing agreement between the Ministry of Justice, The Department for Work and Pensions and Her Majesty's Revenue and Customs, which took two years to establish. See https://www.gov.uk/government/statistics/experimental-statistics-from-the-2013-moj-dwp-hmrc-data-share.

[12] Department for Business, Innovation and Skills (2013). *Improving Access for Research and Policy: the Government Response to the Report of the Administrative Data Taskforce.* London: Department for Business, Innovation and Skills.

[13] Elias, P. (2013) Administrative Data Taskforce: Report of the Technical Group. Swindon: Economic and Social Research Council.

[14] The Taskforce Report uses an example to make this point – the development of the Scottish Longitudinal Study (Department for Business, Innovation and Skills, 2013, p. 10)

[15] UKSA (2014) Administrative Data Research Network Board. Strategic Issues for the ADRN Board and forward planning. Paper ADRN (14) 03.

[16] Cabinet Office (2015) 'Conclusions of civil society and public sector policy discussions on data use in government'.

[17] See https://www.statisticsauthority.gov.uk/national-statistician/national-statisticians-data-ethics-advisory-committee/

[18] Reasons for rejection are: insufficient potential public benefit; (1 case); insufficient scientific merit (2 cases).

[19] See http://webarchive.nationalarchives.gov.uk/+/www.dh.gov.uk/en/Managingyourorganisation/Gatewayreviews/DH_121642 for information about the Gateway Review process. The review team focusses on the progress of major projects, not their scientific achievements.

[20] UK Statistics Authority, Administrative Data Research Network Board, Minute of the Tenth Meeting of the Board, 12 July 2016.

[21] Commissioners for Revenue and Customs Act 2005.

[22] There is an arcane distinction here between statutory and non-statutory government bodies. The former are established by an Act of Parliament. The latter are vestiges of the Crown and reflect the historical ministries set up by the advisors to a king or queen. They continue to exist to this day via what is termed the 'royal prerogative' and have all the rights that a private citizen has, including the right to share data with

others subject to the restrictions placed on private citizens. This right is termed a 'permissive right' in that the government body is not obliged to release data, but may do so in accordance with the legal restrictions covering the processing of personal data.

[23] In collaboration with the Office for National Statistics, the ESRC funded an enquiry into the public acceptability of establishing the ADRN (see http://www.esrc.ac.uk/public-engagement/public-dialogues/public-dialogues-on-using-administrative-data/)

[24] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Despite the fact that the UK will leave the EU, it is highly likely that this regulation will be transposed into British legislation.

[25] This is the direct investment cost from the UK science ministry to the Economic and Social Research Council over the period October 2013 - September 2018, to support the capital spend and staff salaries at a number of UK higher education institutions, as well as the funds paid over to the three statistical authorities of the UK and the UK Statistical Authority for their roles in supporting the network.

[26] See, for example, https://adrn.ac.uk/for-the-public/about-us/