Original Research Article

BIG DATA & SOCIETY

Big Data & Society
July–December 2018: 1–17

DOI: 10.1177/2053951718819560
journals.sagepub.com/home/bds
SAGE

The role and nature of consent in government administrative data

Anna Sexton[1], Elizabeth Shepherd[2], Oliver Duke-Williams[2] and Alexandra Eveleigh[3]

Abstract

This article draws on research undertaken by the authors as part of the Administrative Data Research Centre in England (ADRC-E). Between 2014 and 2017, we conducted four case studies on government administrative data for education, transport, energy and health. The purpose of the research was to examine stakeholder perspectives about the sharing, linking and re-use (secondary use) of government administrative data. In relation to the role and nature of consent given by data subjects for re-use, our study revealed significant variations in data provider and researcher attitudes. Although our study setting was England, we believe that the findings have wider resonance. Our analysis identified six factors which might account for the variations around consent: the specificities of the legislative framework governing the collection and processing of particular data; the type of data being collected and the relational context in which it is created; the broader information governance framework in which the data resides; the creating organization's approach to data release; the relative levels of risk aversity within the creating organization; and public perceptions and social attitudes. In conclusion, we consider whether consent is still the best mechanism available for data re-use, or whether a social contract model of data sharing should be developed.

Keywords

Administrative data, consent, privacy, information governance, data protection, open data

Introduction

Obtaining the consent of the data subject is the primary mechanism which underpins the fair and lawful processing of personal data in a European setting. It acts as a governance mechanism 'to protect individual interests, to promote personal autonomy, and to act as a foundation for trust' (Laurie and Postan, 2013: 372–373). However, consent is not a simple concept and is not easily defined, obtained, or observed. In the age of Big Data, there are questions over whether it is a sufficient mechanism to protect privacy, and the extent to which it can address a wider set of concerns around the 'fairness, justice, and due process' of data use and re-use (Barocas and Nissenbaum, 2014a). In an exploration of the limits of consent, O'Neill (2003) suggests that the inherently propositional nature of consent creates degrees of opacity around what is being consented to. Does consent represent informed choice, particularly when possible outcomes are complex or when the freedoms of the individual may be compromised? O'Neill (2003) also argues that the consent of individuals is not sufficient when the resulting action has implications at a collective as well as an individual level.

Research 'in the public interest' (which uses pseudonymised or identifiable person-level data) may reduce the obligation to obtain consent. In place of explicit consent, researchers and data providers rely on legal gateways, on privacy notices, and on offering opt-outs

[1]Department of Information Studies, UCL, UK; The National Archives, UK
[2]Department of Information Studies, UCL, UK
[3]Department of Information Studies, UCL, UK; The Wellcome Library, UK

Corresponding author:
Elizabeth Shepherd, Department of Information Studies, UCL, Gower Street, London WC1E 6BT, UK.
Email: e.shepherd@ucl.ac.uk

to data subjects. Our research uncovered distinct differences in the role and nature of consent for the research use of government administrative data in England across education, energy, transport, and health. Our analysis identified six factors which might help to explain the variations: (1) the specificities of the legislative framework governing the collection and processing of particular data; (2) the type of data being collected and the relational context in which it is created; (3) the broader information governance framework in which the data resides; (4) the creating organization's approach to data release; (5) the relative levels of risk aversity within the creating organization; and (6) public perceptions and social attitudes. Although our study setting was England, we believe that the findings have wider resonance.

In this article, following a discussion of the qualitative, inductive research methods used, we present a discussion about the data relating to consent from the four case studies in turn (education, transport, energy, and health), followed by a summary of the six factors which influence the understanding of consent identified from the research data. We explore whether a single approach across government to consent for the secondary research use of personal data exists or can be developed or, perhaps preferably, whether a multi-faceted model can be envisaged which would allow for both individual and collective goods and would promote public understanding and trust in government data use. In conclusion, we consider whether consent is still the best mechanism available for data re-use, or whether an alternative social contract model of data sharing should be considered. This research contributes to public policy debates and suggests an approach to the reconfiguration of the social understandings around consent for secondary data use.

In European law governing the processing of personal data, obtaining the consent of the data subject is the primary procedural mechanism which underpins the fair and lawful processing of personal data (General Data Protection Regulation (GDPR), 2016). The UK Information Commissioner's Office suggests that explicit consent 'requires a very clear and specific statement of consent' that is 'expressly confirmed in words' (ICO, 2017). However, according to McDermott (2017), data protection as a fundamental human right is linked to upholding the principle of autonomy alongside privacy, transparency and non-discrimination. This points to an information ecology in which consent may be the *primary*, but is by no means the *only* mechanism through which processing of personal data can be lawful.

Secondary data processing in a health context is subject to the common law duty of confidence, as Grace and Taylor (2013) suggest, which prevents data being disclosed for purposes outside the 'reasonable expectations' of the data subject. In a complex data environment, such as in the NHS, defining public expectations for control of patient-identifiable data and information governance systems is also complex (Caldicott, 1997). While obtaining consent from the data subject for secondary use of personal data is the default position (Grace and Taylor, 2013), the use of de-identified data and of identifiable data in the public interest reduces the obligation to obtain consent. As Barocas and Nissenbaum (2014a, 2014b) note, informed consent has 'longstanding operational challenges' which are exacerbated in an era of Big Data and complex data flows. Anonymised data also has significant limits in an inter-connected data world.

One approach to managing the mutable nature of consent is through 'opt-outs' through which data subjects can object to the use of data for secondary purposes. Precisely defining the opt-out in an understandable way in a form which can be operationalised is not easy, and if done badly, can result in a loss of public trust and the fracturing of the social licence to collect and process data, as exemplified in the failed care.data scheme (Carter et al., 2015; Ostherr et al., 2017; Vezyridis and Timmons, 2017). Sexton et al. (2017), MacNeil (2011), Flinn and Shepherd (2011), and others have explored the trust nexus surrounding records and data, suggesting that public trust is a fluid construct, dependent on an ever-changing dynamic between data, the space-time in which it exists, and those that interact with it.

Taylor (2015) argues that the way forward through the dichotomy of opposing tensions is to ensure that data protection is framed not just in terms of individual rights and freedoms but as a collective public good. A number of authors have examined issues around the focus of consent on the autonomous individual, the binary choices presented by consent to privacy notices as a condition of accepting a service (Cate and Mayer-Schönberger, 2013), and the increasingly social context of consent, transformed by digital data (Ruppert et al., 2013). Social contract theory from the work of moral and political philosophers, including, in the 20th century, John Rawls (1958), might help us to rethink consent in a social context. A social contract theory seeks to show that citizens 'comply with the fundamental social rules, laws, institutions, and/or principles of that society', 'by rational agreement' even though individual reasons for complying differ (Stanford Encyclopedia of Philosophy, 2017). Traditionally, the social contract relied on consent, but contemporary social contract theory has moved towards agreement and the question of justification (Rawls, 1958). This aims to model the reasons and conditions under which citizens would agree, if they were properly

informed about an issue and acted in a reasonable manner.

These ideas change the frame of the discussion on privacy, rights, and data protection away from a focus on individual data subject rights towards achieving collectively beneficial outcomes. It may be possible to remodel the role that consent should play in balancing individual privacy and wider societal benefits through a social contract model for data sharing.

Research methods: Four instrumental case studies

This article draws on research undertaken by the authors as part of the Administrative Data Research Centre in England (ADRC-E). Between 2014 and 2017, we conducted four case studies on government administrative data for education, transport, energy, and health. The purpose of the research was to examine stakeholder perspectives about the sharing, linking, and re-use (secondary use) of administrative data. We chose the four cases as major central government policy areas which make extensive use of administrative data, that would provide a sufficiently extensive range of data providers and datasets, offered variable levels of public awareness of secondary use of data (higher in education and health, than in energy and transport), and sufficient numbers of academic researchers using administrative data to enable the research. The qualitative study undertook 44 semi-structured interviews, plus one focus group of 4, as the main data collection method. Interview data was triangulated with documentary analysis (such as data access protocols, consent forms, etc.) and set into a literature review. The secondary use of government administrative data by academic researchers was the central focus through which the interviews examined data and stakeholder issues. Gathering perspectives from academic researchers (at a variety of career stages), who use government administrative data in their research, therefore formed the core of each case study, and their perspective is most fully reflected in the results. However, interviews with other stakeholders including government bodies acting as data providers, policy makers, advisors, regulatory bodies, research funders, and lobby groups enriched the data. In the education case study, we also interviewed five data subjects. The stakeholders represented in the research are more fully explored in Sexton et al. (2017). A summary of the interviewees is given in Table 1. An example semi-structured interview protocol which framed the questions for the academic researchers is included in Table 2. All interviewees were anonymised, and extracts used in this article are attributed to individual interviews by use of a code (e.g. A10). We articulated the study's research questions through three themes which framed the data analysis: trust, risk, and consent. Interview transcripts were thematically coded line-by-line, captured in NVivo 10, assigning a coding label to each component and refining the codes into themes derived inductively from the data, in an iterative process of analysis and assigning meaning. Codes were reviewed, at the level of the coded data extracts, and later in relation to the entire set of data, to ensure the themes reflected the participants' unique perspectives grounded in the data. This article concentrates on our analysis of *the role and nature of consent* in government administrative data.

In the education case study, the main dataset used by researchers was the National Pupil Database (NPD); a person-level database which matches pupil and school

Table 1. Summary of interviewees in ADRC-E study.

Case study	Number of interviewees	Types of interviewees	Coding range	Dates of data collection
Scoping study	5	Academic researchers	A1-A5	June 2014–July 2014
Education data	12, plus 4 in a focus group	University students (data subjects), HEI student data manager, academic researchers, data providers (DfE)	FG1, A6–A17	December 2014–April 2015
Transport data	9	Academic researchers	A22–A25, A28, A29, A31–A33	September 2015–November 2015
Energy data	5	Academic researchers, data providers	A26, A27, A30, A34, A35	October 2015–January 2016
Health data	18	Academic researchers, data providers, policy advisors, information and data managers, data subject representatives	A36–A53	May 2016–October 2016

HEI: Higher Education Institution.

Table 2. Example interview protocol.

ADRC-E Academic Researcher Interview Protocol

Introduction

The interview questions have been structured according to the five phases of the Data Documentation Initiative's research data lifecycle model (DDI lifecycle http://www.ddialliance.org). You will probably have a more detailed knowledge of some phases of the model than others, and it may help to keep one or two specific examples in mind as we work through the interview.

- Can you give me a brief outline of the type of research that you do, including any relevant recent projects which have made use of government administrative data?
- Does your research involve purely administrative data, or do you link administrative data to other data sources (e.g. survey data, longitudinal cohort studies)?
- Does your research use administrative data from a single government source, or from two or more departments? Please give details.
- Do you use this data with a single research purpose in mind, or do you (intend to) re-use the same dataset to investigate other research questions?
- Do you know whether the data you use in your research is also used by other researchers and/or for other use purposes? If yes, please give details.

Discovery and planning

Designing research

- Can you describe your first steps in planning to use government administrative data? Who do you approach initially and what level of detail do you provide to them about your proposed research? i.e. do you have a defined idea of which variables and/or datasets might be useful, or do you start out with a more general idea that the data provider might be able to supply datasets of interest?
- If you were advising a researcher in your field who was hoping to use government administrative data for the first time in their research, what tips would you give them on how to go about applying for access to such data?
- Do you review other researchers' work using the same or similar data before you approach the data provider for access for your own work?
- At what point in the research design process do you first get in touch with the prospective data provider? How much lead time do you need to build into the planning process?
- What ethical approval processes are you required to undergo at your home institution? To what extent does the data access application procedure for government administrative data duplicate or build upon this standard ethics procedure?

Planning data management

- What kinds of guarantees do you provide to the data provider about how you will manage the data, and access to it, during the course of your research project?
- Are you under any obligation to (a) your home institution (b) your funder(s) (c) your collaborators' institutions or any other third party concerning data management and access? Please give details. Are there any conflicting requirements and how are these resolved in practice?

Planning consent for sharing

- Do you typically apply for access to government administrative data for yourself alone, or for all the members of your research group, or for yourself and named collaborators?
- Are any of these collaborators based outside of the European Union?
- What personal details and level of detail about the aims, methods and outputs of your proposed research are you asked/do you expect to give to the data provider before your application can proceed? Do you know how the data provider processes this information and what checks are made?
- Do you propose to share access to the data with anyone outside of your research group or institution?
- How long does it typically take from application to being granted approved researcher status?

Planning data collection, process protocols and templates

- Can you provide examples of any application forms or protocols supplied by the data provider(s)?
- Does the data provider give any support in putting together a formal access request?
- Can you provide examples of data access or permission agreements, confidentiality protocols, approved researcher application forms, etc. that you have in place for current or previous research projects?
- Do you ask, and is it possible, to view samples of the variables and data you are hoping to gain access to, in order to help plan your research?

Finding and discovering existing data sources

- What documentation is available to you about the data content itself? How much do you typically know about the circumstances in which the data was originally collected by the data provider? e.g. do you have examples of forms used to collect the data? Are you aware of any confidentiality promises made to data subjects at the point of collection?
- How do you submit your application for access?
- How clear are (a) the application process (and any guidelines) (b) the approval criteria, to you as a researcher?

(continued)

Table 2. Continued

ADRC-E Academic Researcher Interview Protocol
• Who is responsible for reviewing the applications and for granting approvals? • How long does it typically take from submitting your application to approval being granted for release of the data? ***Data collection*** **Collecting data – recording, observation, measurement, experimentation and simulation** • What forms of filtering are performed on the data before it is released to you? Do you think this filtering is reasonable or excessive? Are you able to request customised preparation of the data? e.g. pseudonymisation, customised data release, removal of geographic identifiers, etc. • What kinds of data output are you looking for? e.g. aggregated datasets, cross tabulations, regression coefficients **Capturing and creating metadata** • Typically, what kinds of supporting information (metadata) are available about the data itself? e.g. codebooks **Acquiring existing third party data** • Is any charge made for access to this government administrative data? • Are there any other access or redistribution restrictions? ***Data processing and analysis*** • What physical security features are in place to prevent unauthorised access to government administrative data at the location where you carry out your data analysis? e.g. electronic card entry, CCTV, webcams, secure room, etc. • What technological infrastructure is used to provide access to (and prevent unauthorised access) to the data? e.g. VPN, Citrix, etc. • Are you required to validate your identity on each occasion that you wish to use government administrative data, and if so, in what way(s)? e.g. username + password, smart card technology, biometrics • Are you required to undergo any form of training before being granted access to government administrative data? If yes, can you describe what this training covers and how it is delivered (face-to-face, online, etc.)? Do you think this training is an adequate preparation for working with government administrative data? • How far are you able to conform to the terms of the data provider's licence granting you access to data for research? Are there any particular stipulations which you have difficulty in meeting? i.e. what is the gap between a strict interpretation of the data provider's terms and actual research settings in practice? **Entering data, digitising, transcribing and translating** • What software do you use to analyse the data? e.g. SPSS, STATA, etc. **Checking, validating, cleaning and anonymising data where necessary** • Do you check or validate the data provided to you in any way before beginning your research analysis? • Do you perform any further cleaning or anonymisation of the data (beyond any filtering carried out by the data provider)? • What impact would you say filtering of data (whether performed by the data provider or by the researcher) has upon (a) the feasibility of different kinds of research question you would like to explore (b) the validity of your research? **Deriving data** • Are you permitted to create new datasets by combining data from one source with another dataset from the same source or another provider's dataset(s)? If yes, are there any limitations on the kinds of data linking you are permitted to carry out? • Would you like to compare government administrative data from England with the same or similar data relating to (a) other countries within the UK (b) other countries within the EU (c) countries outside of the EU? Is this data use permitted? **Describing and documenting data** • Do you document the process(es) and methods you use to prepare or analyse the data? Can you provide examples? **Analysing data** • What support is available from (a) the data provider (b) your institution in regard to analysing government administrative data? e.g. running researcher-supplied code • Would you like any further form of support which is not provided? Please give details. **Interpreting data** • Does the data provider offer any support in understanding, analysing or interpreting their data? If yes, can you describe what form this takes? • Do you find you need to go back to the data provider to ask for further information on the data source(s) in the light of the findings emerging from your research? **Producing research outputs** • What criteria are put in place for disclosure review before publication or other dissemination of research outputs?

(continued)

Table 2. Continued

ADRC-E Academic Researcher Interview Protocol
• Is any charge made by the data provider for disclosure review? • Does the data provider insist upon a right of review before publication or dissemination of research outputs? If yes, how is this done and how long does it take (review of draft analysis, final text, lead-in time allowed, etc.)? **Authoring publications** • Does the data provider seek to be named as a co-author in publications or other public dissemination of the research? • How does the data provider require their input in providing the data for the research to be acknowledged? **Citing data sources** • How are data providers and data sources acknowledged in publications and other research outputs? **Managing and storing data** • Who is responsible for providing storage for government administrative data during the course of your research (researcher, researcher's institution, data provider, third party provider, e.g., UKDA) and is there any charge associated with this storage? • What security features (physical and/or technological) do you put in place to safeguard access to preliminary and interim results of your analysis based upon government administrative data? • Are these locations accessible to anyone other than the approved researchers for each specific project? ***Publishing and sharing*** **Establishing copyright of data** • Does the data provider seek to obtain any rights over the publication or dissemination of research based upon 'their' data? **Creating discovery metadata and user documentation** • Do you create any metadata or documentation about the dataset(s) in the course of your research? If yes, do you offer this information back to the data provider for the benefit of future users? • Do you have any ethical or licence-based duty to report errors in the dataset back to the data provider? • Does your research produce machine-readable data? If so, what formats and procedures are used to do this (RDF, etc.)? Is this data offered back to the data provider or shared more widely? **Publishing or sharing data** • What technological infrastructure is in place to share preliminary research results with the data provider? e.g. VPN, encryption • What statistical controls are put in place in research outputs to prevent accidental disclosure? e.g. aggregated results • Is formal permission or clearance required from the data provider before proceeding to publication? Do you have examples? • What expectations (obligations or commitments) are there from (a) your home institution (b) funder(s) (c) collaborators' institutions or other third parties in terms of publication and open access to publicly funded research? Do these requirements conflict in any way with the data provider's licence terms? If so, how to do you manage the conflict between the two and how onerous do you find this responsibility? **Distributing data** • Are you free to distribute your results multiple times without resubmitting a publication clearance application to the data provider? **Controlling access to data** • Are you aware of any unauthorised disclosures – accidental or deliberate – of government administrative data within your area of expertise? • Is the approval procedure a one-off, or is there a mechanism for re-assessing your data access request during the course of your research project, or as other relevant data comes to light? • Is the research approval process a one-off, or do you have to re-apply for each new dataset you wish to use? **Promoting data** • In what ways is it possible for peer-reviewers to establish the validity and reliability of your analysis? ***Long-term management*** **Preserving and curating data** • What are your obligations or commitments to (a) your home institution (b) funder(s) (c) collaborators' institutions and other third parties as regards the long-term retention of the datasets upon which your research is based? • Are you aware of any formal data management commitments made in the original proposal for funding? Can you provide copies of these requirements? • Do you have access to the technological and administrative infrastructure required for the long-term storage and management of datasets? Is this (a) an in-house institutional service (b) a service offered by the data provider (c) a third party supplier? Please give details of any such service you have used in the past or might consider in the future. **Migrating data to best format** • Do you transform the data into other formats in the course of your research? (.csv, .sav, .dta, etc.) Please give details.

(continued)

Table 2. Continued

ADRC-E Academic Researcher Interview Protocol
Migrating data to suitable medium
• Do you have a data management plan to cover hardware/software replacement where you require local access to data over the term of a multi-year project?
Backing up and storing data
• What happens to the data you have used after publication or when your research project funding ceases?
• What happens to your preliminary or interim analysis results after publication or when your research project funding ceases?
Gathering and producing metadata and documentation
• (If interviewee indicates in-house or third party supplier of long-term preservation infrastructure) What supporting documentation (metadata) about the dataset is handed over for long-term retention and how is this linked to the dataset?
Reusing data
Conducting subsequent analysis
• Do you ever undertake further or subsequent analyses using the same underlying dataset(s)? If yes, please give an example.
Undertaking follow-up research
• How straightforward would it be to undertake follow-up research based upon the same or similar data?
Conducting research reviews
• Are meta-analyses commonly conducted in your field? Would it be possible to include your results in such meta-analytical studies? What difficulties might a researcher wishing to include your administrative dataset studies in a meta-analysis encounter?
Scrutinising findings
• What level of access would be required in order to attempt to replicate your results?
Using data for teaching and learning
• Do you use government administrative data in your teaching at all? Please provide examples. What difficulties do you/might you encounter in using such data in your teaching?

characteristic data to pupil attainment. Access requests are scrutinised by the Department for Education (DfE)'s Data and Education Standards Analysis Group and the Data Management Advisory Panel (DMAP) (UK DfE, 2016). Researchers can apply for data linked to Higher Education Statistics, and to Further Education students' Individual Learning Record.

In the transport case study, many researchers relied on open data. The STATS19 database of road traffic accidents resulting in personal injury, available from the Department for Transport (DfT), is collected by the police, validated by local authorities, and collated at the DfT. The dataset, published annually, is available through the UK Data Service, as open data through data.gov.uk, and directly from the DfT.

The key dataset in the energy case study was the anonymised dataset of the National Energy Efficiency Data (NEED) Framework, collated by the Department of Energy and Climate Change (DECC). It brings together data on gas and electricity consumption provided by UK utility companies, Home Energy Efficiency Database, property values and household characteristics from other government agencies, utility regulators and external data providers (including commercial credit brokers). The data in NEED is 'publically available data and data provided through commercial licences, voluntary agreements and service level agreements with dataset owners' (UK Department for BEIS, 2013).

The main dataset that researchers in the health case study used was Hospital Episodes Statistics (HES); a patient-level database containing over a billion records of patients attending Accident and Emergency units, admitted for treatment or attending outpatients clinics at National Health Service (NHS) hospitals in England. NHS Digital publishes annual HES data from 2009 at provider-level as open data. Researchers can access anonymised in context (or more fully identifiable), patient-level data. Access requests go to the Data Access Request Service of NHS Digital, and through IGARD (the Independent Group Advising (NHS Digital) on the Release of Data), replacing the Data Access Advisory Group since 2017.

Results: Four case studies

This section examines the data from each of the four case studies in turn, drawing out the findings around the role and nature of consent in government administrative data.

The role of consent in enabling research use of routinely collected administrative data for education

In recent years, the DfE has been at the forefront of government data sharing initiatives, leading in an 'exemplary way' (A12) on ensuring that routinely

collected government education data can be used 'extensively by researchers' (A12). This open attitude to sharing data for secondary research use has been part of a cultural shift where a more risk averse attitude of 'not giving the data to anyone unless they had to' (A16) has been superseded by a deeper dedication to 'widening access' (A13). As one respondent commented:

> You know, some people think we should just put the National Pupil Database on the web so anyone who wants to use it… and obviously in a world of open policy-making and open data, there's much to be said for that. (A16)

This attitudinal shift in DfE reflects changes to the regulations governing the sharing of individual pupil data, where a narrow definition of acceptable research was broadened to legitimise use of individual pupil data for the social and collective 'purpose of promoting the education or well-being of children in England' and 'conducting research or analysis, producing statistics, or providing information, advice or guidance' (UK Government, 2013). Our interviewees saw this as a positive impact on moving 'the research field forward' (A12). The DfE also introduced more transparent, well publicised and consistent access protocols to its datasets, notably the NPD. More researchers are accessing and making use of education data than has previously been possible, for a much wider range of research initiatives.

The DfE's open attitude to secondary use can be set against the relatively peripheral role that consent plays in enabling data collection. In the NPD, school data collection is mainly mandatory under specific legislation and regulation (Education Act 2005 section 114, Education Act 1996 section 537A, Children Act 1989 section 83, UK Government, 2013), and relies upon the display of privacy notices in schools and on local authority websites. A limited opt-out arrangement is offered (to parents of children aged 13 upwards to their 16th birthday, and thereafter to the students themselves) which restricts the exchange of data with local youth support services, but there is no equivalent opt-out mechanism in respect of the same data passed by schools to local authorities and to the DfE. As described by A13:

> Privacy notice[s] tell parents, teachers and children why we need to collect the information, what we're allowed to use it for, and that covers any request that we will get that we will then approve, so we don't have to go through actual permissions from the schools [or individual pupils] themselves…so we have the legal side covered to allow us to release this data.

Secondary research use of pupil data held in the NPD is legitimised without an individual data subject's consent. Pupils and their legal guardians seem generally tolerant of the lack of provision for consent and opt-out: one parent's blogpost response to the DfE consultation objecting to the lack of consent provision is exceptional (infiniteideasmachine, 2012).

Governance of requests to re-use education data held by DfE for research is provided by the DMAP which includes external representation (UK DfE, 2016). Data in NPD is broken down into four tiers. Tier One data are directly identifying (includes names, addresses, date of birth, exam candidate number, unique learner number) and/or highly sensitive (includes looked-after status, ethnicity, Special Educational Needs assessments, reasons for exclusion, and absence). Requests to use this data always go through DMAP, as do all requests for linkages with other datasets. Tiers Two to Four relate to less identifiable and sensitive classes of data, which are usually handled by the NPD and Data Sharing Team. Governance oversight seeks to ensure that data requests are in line with legislation and regulations, whether the data requested is proportionate to need (data minimisation) and security standards are proportionate. For highly sensitive and identifiable data releases, a finite period of time for which access is permitted for research is agreed under specific licensing agreements. The consent of data subjects, or their legal guardians, for such secondary re-use is not required.

However, where researchers seek to link independently collected cohort data to routinely collected administrative education data (such as the NPD), the consent of data subjects *is* required as the independently collected data is not covered by the legal gateways. Here, in keeping with a proactive approach to data sharing, DfE governance is light touch:

> If a researcher has collected some data outside that area and wants to link it to the NPD…they have to show us that they've got the relevant consent arrangements in place. We don't need to see evidence of that consent anymore, but the onus is on them. (A13)

Some of our interviewees questioned whether it is ethical to use individuals' data for research or data linkage without consent, and whether the privacy notices were sufficiently detailed about secondary uses. A8 remarked: 'I really buy into the idea that people should be told that if their data's going to be used for a secondary purpose that that needs to be made clear to them at the time that they provide that information'.

A12 argued that asking for individual consent for research purposes is problematic not only because of

its likely impact upon research re-use of the data but also because it is likely to cause problems for existing government uses of data for monitoring educational progress. A12 set out the difficulties:

> If you ever made it optional, the government would stop functioning because, obviously the administrative data is not there for research, the administrative data is there first and foremost for them to run the system, so if you started to have to ask every individual whether or not they can put their data in this database, um, well, you couldn't do that [laughs]!

Students (who were data subjects) had different views on whether it was reasonable to gain consent from them. One expressed anxiety about a perceived lack of control over how data was used, even if they did not object to their data being used for research:

> I think my biggest issue is that I don't know what I have and haven't agreed to. . . .I don't mind sharing it because I think there's a use to sharing this, and pooling it together for research and analysis, but it just worries me that I can't turn round and say I know exactly what sort of information is being held about me, how, and for what reasons. (A20)

While another reflected that the system did not allow for changes over time: '. . . as a child at school, you may have had data collected about you, and then, later in life, you know, you may encounter certain issues and want to retract it' (A18).

In summary, in relation to research use of routinely collected education data such as NPD, for the most part, the collection, sharing and linking of individual-level personal education data for research use are unconsented. Research use of government administrative data for education is subject to a legislative and regulatory framework which enables secondary use largely without consent, relying instead on a system of privacy notices. Where consent is deemed necessary for specific uses, the DfE takes a 'light touch' approach with researchers responsible for obtaining consent. Although some of the data collected is considered sensitive (Tier One), most is not, and the data is collected in a mandatory context, with a limited system of opt-outs for data sharing. DfE operates a clear governance framework which generally facilitates data release for research use, suggesting a relatively low level of risk aversity with respect to data release. Public perception seems surprisingly tolerant of the lack of explicit consent mechanisms for NPD, when compared, say, with health data, although students expressed some concerns about their lack of control over their educational data.

Transport data and open data

In relation to government administrative transport data, the role of data subject consent in determining research re-use of the data is also peripheral, but for entirely different reasons. The Department for Transport (DfT) has been at the vanguard of efforts to facilitate better use of, and access to, data for more efficient government and the public benefit. DfT commissioned research into a UK National Transport Data Framework (NTDF) (Landshoff and Polak, 2008), in advance of the government-wide push towards open government data and, in the wake of the Shakespeare Review (2013) which considered the growing value of public sector information and how it could be better exploited including recommending the publication of a National Data Strategy, published an open data strategy (DfT, 2013). The strategy states that 'transport is a 'data-rich' area, where there is huge public appetite for information that can be used to inform travel choices, to improve performance and to hold operators and Government to account', including 'datasets owned and published by the wider industry such as timetables and real-time running' information. Movement towards open data by DfT and its agency family (including the Driver and Vehicle Licencing Agency (DVLA)) resulted in public availability of large numbers of official statistical datasets relating to road and public transport. DVLA has been a leader in developing digital government services, enabling sharing of data with the motor trade and insurance industries, local government and the police, as well as launching online access for individuals to their own driving licence records (https://www.gov.uk/view-driving-licence). The sale of DVLA vehicle register data, although specifically permitted under statute (UK Government, 2002), has not been without controversy. DfT's open data strategy focused on making data available that is of use to service providers, application developers, and to individual citizens as data users (DfT, 2013).

Academic researchers interested in transport issues therefore mainly have to rely on anonymised aggregated open datasets (such as STATS19), rather than having access to the more granular person-level administrative data from which these open datasets are derived. As an aggregated and anonymised dataset, individual consent from data subjects in STATS19 is not required for governance. Our interviewees acknowledged that injured individuals' consent was unlikely to have been obtained by police attending the scene of an accident, but this was contrasted by A24 with health data (potentially relating to the same incident) where the expectation around obtaining consent was the opposite.

Given that much data in the transport arena has long been open and publicly accessible, it was a common view that no explicit consent or ethical review was required for secondary research use, of STATS19 for example:

> Generally speaking when it's secondary analysis of data that's publicly available you do not [gain] ethics approval, and you know, as long as I think no one is identifiable which they're not because, you know, it's based on collision location, that's sort of the only variable that could give away anything. (A22)

Some interviewees additionally expressed a view that research use in the public interest would be a 'reasonable expectation' of individuals submitting data to the government:

> I suppose my slightly hardline view, ... if people are ... giving data to say the government, then I think it's a reasonable expectation that that data will be used for, research or whatever it is for the betterment of the way things are run. (A28)

And there was a broad agreement amongst several interviewees on the impracticality or illogicality of obtaining post hoc consent for data linkage:

> I think that if someone was doing some research that was going to be for the public good that required linking several datasets, I think it's crazy to say that can't go ahead unless you've got the approval of all the people who are in the dataset. I think the consent side will stifle some good research ideas and will hold back some research. (A24)

The nature of the transport industry and the widespread re-use and linking of aggregated and anonymised transport data for commercial and traveller benefit, combined with a general lack of academic researcher access to personal and identifiable data, result in data subject consent being considered a minor issue by researchers and data providers. Public perception seems not to have affected this stance, in spite of controversy over the sharing of DVLA data (BBC News, 2012).

Energy data and multiple data sources

A shifting landscape around consent emerges in our third case study on government administrative data for energy. DECC collates the NEED framework which derives from multiple data sources and does not solely comprise government administrative data. Instead, multiple data flows of varying provenance are linked together in-house by DECC, for the purposes of generating government statistics and for internal research purposes. The statutory basis for DECC's acquisition of the data at the heart of the NEED framework – energy consumption data derived from utility meter readings – lies in the Statistics of Trade Act 1947, as amended by subsequent legislation, such as the Electricity Act 1989 and the Utilities Act 2000. Unsurprisingly, given that the underpinning legislation is 70 years old, it does not anticipate contemporary data uses. This legislative framework, coupled with data licensing restrictions and reinforced by the effects of privatisation in the utilities sector, severely limits the extent to which DECC is able to share NEED data beyond government:

> ... the Statistics of Trade Act, and that allows the government to use, for statistical and research purposes, information from businesses which is considered to be useful to the government. ... we cannot disclose that information in such a way that any individual business or any individual entity within the data can be identified. So we can only publish aggregated figures and we can't pass the data on without the express permission from the people that have supplied the information. (A26)

The restrictions in the legislation mean that identifiable data cannot be disclosed beyond government, and even the release of aggregated and anonymised data requires provider level agreement. Despite these legislative barriers, DECC has made conscientious efforts to encourage research use of the data by gathering consent from providers and making an anonymised, partial dataset available through the UK data service:

> So earlier in the year, we wrote to all of the electricity companies and all of the gas companies, and asked their permission for us to share their data with academic research partners that we thought put forward proposals which were worthwhile. We put lots of caveats around everything to say that we'd ensure that all the appropriate safeguards were put in place and all the appropriate protocols were followed to ensure the data is held securely. But unfortunately, a number of the suppliers declined, and said they weren't happy for that to happen. (A26)

Academic researchers are forced to bypass DECC and seek access to the individual elements of the complex data flows which make up NEED (see for example Critchley et al., 2007; Foulds and Powell, 2014). An alternative mechanism is for DECC to commission external researchers in order to provide privileged access to specific academics. Even then, DECC will

'still have written to the companies to tell them that that's happening, we give them the opportunity to object, it's kind of like an opt-out.' (A26)

In place of consent by individual householders, NEED relies on a privacy impact assessment (PIA) in conjunction with the energy suppliers and the regulator to address transparency concerns. It requires that 'all suppliers have a privacy policy which provides information on how customers' data is handled', including a fair processing notice on how data may be shared with government. The PIA allows government to link data 'in circumstances where permission has been granted by the individual (e.g. through a survey response) or an appropriate privacy notice has been published', but does not enable access by external academics or other third parties: '. . .if we wanted to link the data in-house, that's fine . . .. But in order for an academic to link to this data, we'd have to give them that data, and that's the bit we can't do' (A26).

However, the centrality of the utility companies in consenting and approving data access and re-use is set to change. The presumption that the utility companies own the data and therefore are in a position to dictate its flow is being challenged by government policy around the introduction of 'smart meters'. While the utility company owns the meter:

> Ownership over the data that you use in your business and your home is becoming more personalised and it has been decided under government processes that you, the homeowners or the person who controls the meter will own that data, and that the suppliers will not own that data at all. And you will be the one that grants access to it. (A27)

The national roll out of smart meters has therefore brought with it a significant change where research use of energy data will require customer (data subject) consent: 'smart meter[s] . . .have put the energy consumer, householder in control of the data. So any use of the data for research purposes must gain specific consent' (A30).

In the energy data case study, the legislative and regulatory framework surrounding NEED has subsumed the role of consent to a minor position. The complexity of the relational context in which utility data is collected and of the information governance frameworks is increased by the mix of public, private, and third parties who deliver energy services. If consent is required for secondary research, it has to be given by the data owners, currently often a commercial utility company: however, in future, the use of smart meters to collect household data will transfer data ownership to individuals and that will have significant implications for the need to gain consent for secondary research use of energy data. No doubt, a programme of public information about consent and any opt-outs around energy data and the public interest in research will have to be developed.

Personal sensitive health data and the role of consent

Across our four case studies, the role of consent as a mechanism for enabling secondary uses of data is arguably most complex in the context of government administrative health data. In addition to the broad legal framework, disclosures that have a 'robust public interest' and therefore permissible without explicit consent led to the introduction of Health Service (Control of Patient Information) Regulations in 2002 (commonly referred to as 'Section 251'). An independent group, the Confidentiality Advisory Group (CAG), advises the Health Research Authority (HRA) and the Secretary of State for Health on whether to permit processing for both research and non-research purposes without consent under the Section 251 regulations. Two significant safeguards are whether the research is in the 'public interest' and the lack of a 'practicable alternative'.

Analysis of our interview data indicates that the existence of the Section 251 provision is broadly welcomed by the range of stakeholders we interviewed (including health researchers, policy makers and data providers) as a means of providing a mechanism to judge when secondary use of identifiable and sensitive health data without consent is legally permissible. The complexity of the interface between different aspects of the legislative framework can lead to confusion amongst the research community over when an application to CAG may be necessary. However, there was a strongly articulated perspective from a small minority of interviewees that questioned the legitimacy of having a legal override to consent:

> The 251 process colloquially is bust, it's indefensible, it's a sham, because in theory you should only have a 251 exemption if you cannot contact the patient to get their permission. So in what case in the modern world is it not possible to contact the patient to get their permission? (A36)

Grace and Taylor (2013) argue that under the exercise of powers through the Health & Social Care Act 2012, NHS Digital is effectively released from a duty to provide information about its processing in regard to mandated disclosures, and even the more limited right given to data subjects to object to processing under GDPR is effectively curtailed. Despite the lack of a legal requirement to provide mechanisms to uphold patient

objections to mandated disclosures, mechanisms have emerged for doing so in the form of an 'opt-out' model (Taylor and Taylor, 2014). A Type 1 opt-out prevents identifiable information being shared outside of the individual's GP practice for purposes beyond direct care. A Type 2 opt-out prevents NHS Digital from sharing 'personal confidential' information outside of NHS Digital for purposes beyond direct care. It was strongly argued across our interviewees that the wording of the opt-outs lacks clarity, making it difficult for the public to know exactly what they are opting out of, and consequently also for NHS Digital to know how to apply the opt-outs. NHS Digital faced complex technical barriers which prevented it from extracting the opt-out data from GP data, resulting in a period of suspension in 2015–2016, during which the opt-out was put on hold and confidential patient data was shared against patient wishes (HSCIC, 2015, 2016). Considerable confusion was voiced by our interviewees as to the circumstances in which the opt-outs apply to dissemination of data to researchers.

Most research uses of HES data are not affected by the opt-out because person-level data is supplied to researchers pseudonymised (anonymised in context). However, where researchers want to link data, for example, between HES and Office of National Statistics mortality data, the data given to the researcher becomes re-identifiable. The question of whether such research requests are subject to the opt-out has been cause for concern. A39 reported research into child mortality which was affected by Type 2 opt-out, but queried the lack of transparency about the permission pathways, and who is making the decisions on the application of opt-outs, and whether those decisions can be challenged.

A39 goes on to describe the impact that applying the opt-out is likely to have on the validity of the research:

> If applied it will mean that 2% of the population will be opted out, if it is in Manchester it will be up to about 6% and at some GP practices it will be 100% so that is a very important loss of a non-random section of the population that you really need to know about…the opt-out has the potential to undermine the use of an important piece of health information, and ultimately damage research in the public interest.

When a researcher seeks access to identifiable and sensitive data (or linking a dataset held by NHS Digital to identifiable and sensitive cohort data), IGARD requires that either Section 251 supports the proposal to provide a legal gateway for access without consent, or that the researcher has gained the informed consent of every data subject. Whereas the DfE's DMAP takes a light touch in relation to its role in checking informed consents are in place by placing the onus on the researcher, IGARD takes a far more proactive role. Wordings of informed consent forms have been scrutinised by the panel and, in recent cases, rejected on the grounds that insufficient explanation has been given to the data subject to constitute 'informed' consent. A24 gave an example of extensive discussion over several years, requirements to change information sheets mid-trial in hundreds of hospitals, leading him to conclude, 'I think they are so cautious or so risk averse they're not sure what the requirements are.'

Dame Fiona Caldicott conducted a review (2016) of the basis upon which information is shared in health and social care. The review proposed 'a new consent/opt-out model to allow people to opt-out of their personal confidential data being used for purposes beyond their direct care', although 'where there is a mandatory legal requirement for data in place, opt-outs would not apply'. Even if patients elected to opt-out of data sharing for research and service improvements, they could reverse that decision later and could give specific consent to be included in a research project. At the time of writing, these recommendations have not been implemented.

In summary, in relation to health data, consent (and related opt-outs) plays a far more centralised role in governing research access to data than is the case for use of other types of government administrative data. NHS is generally reluctant to release data for secondary purposes except through highly scrutinised and regulated information governance processes, in spite of open data priorities at a national level. The specificities of the legislative gateways for health and social care data are complex, and the consent and opt-out arrangements are difficult to understand or explain clearly to data subjects. The types of data created are typically highly sensitive personal data collected in a confidential setting, often face-to-face with a health professional, which brings high expectations of trusted data systems and which GDPR recognises as requiring stricter processing. Poor public perception of data management within the NHS, following high-profile data breaches, ransomware attacks, and, in particular, the failed data sharing project, care.data, contributes to a risk averse stance towards releasing data for secondary research.

Results: Six factors

Our analysis of the four case studies suggested that the role that consent plays in both enabling and restricting research use of routinely collected government administrative data varies between case study sectors and creating organisations. The factors governing the role of consent in research use of administrative data which emerged from the inductive analysis of our interview data cluster around six issues. The six factors observed

in our data are not evenly distributed across the case studies.

The first factor relates to the *specificities of the legislative framework* governing the collection and processing of the data. The UK Data Protection Act 1998 (succeeded by GDPR in 2018) and the Human Rights Act 1998 uphold the rights of data subjects in the processing of personal data. Within each case study, these overlap with specific legislation, for example, relating to statutory powers of government to mandate data collection, producing a unique context in which consent is framed. For example, the Statistics of Trade Act 1947 governs data collection by DECC and requires consent for further dissemination, but the focus of that consent is on the utility company as the supplier of the data rather than the individual data subject.

Within the legislative framework, the *type of data being collected and the relational context in which it is created* has an impact on the role that consent plays in governing research access. Health data is created as part of a confidential relationship between data subject and health professional which brings the common law duty of confidence into the framework and reinforces consent as a mechanism for secondary data use. Health data was also classed as sensitive under the Data Protection Act 1998 and as a special category in the GDPR and Data Protection Act 2018, including a requirement for 'explicit consent' to be gained before its secondary use. Consents gained by researchers working with health data are scrutinised more closely by NHS Digital, than those obtained for educational data held by DfE.

Thirdly, the *broader information governance framework* in which the data resides also has an impact on consent. A 'consent by default' approach has emerged as the accepted path to data sharing in the NHS. This is reinforced through the NHS Constitution for England (2015) which, in relation to 'consent, respect and confidentiality' states, 'you have the right to be informed about how your information is used' and 'to request that your confidential data is not used beyond your own care and treatment…, and where your wishes cannot be followed to be told the reasons including the legal basis'. Health care data subjects are offered the right to object to further dissemination through 'opt-outs'. Information and data governance frameworks are more fully developed in the NHS than in other parts of the public sector in our study.

The *creating organization's approach* to data release is the fourth factor influencing consent as a governance mechanism for research access. The DfT, for example, has focused its efforts on publishing open data and deflected attention away from the provision of research access to more granular, person-level, data. Consent as a governance mechanism is not needed, but at the cost of having less useful data available for research.

The *relative levels of risk aversity within the creating organization* is the fifth factor. In the context of health, NHS Digital takes considerable measures to scrutinise the 'informed consent' gained by researchers to enable secondary use of identifiable and sensitive data. The risk aversity seems to be related to the public scrutiny applied to NHS Digital to account for its data dissemination practices and improve its levels of oversight and audit. For example, the Partridge Review (2014) considered data releases made by one of its predecessor organisations and recommended that NHS Digital should tighten mechanisms for compliance and accountability in an attempt to entirely eliminate personal data breaches.

Finally, *public perception* plays a part in determining organizational levels of risk aversity, which in turn has an effect on approaches to consent. In the case of health, prominent criticism of data sharing programmes from the media and lobby groups (such as the failed attempt to share health data across hospitals and doctor's practices, care.data, see Carter et al., 2015) led to the introduction of opt-outs for mandated data collections. However, studies by Ipsos MORI (2014, 2016) on the relationship between public understanding and public trust in the uses of data, and by Health e-Research Centre (2016) asking to what extent patients should control access to data, indicated more positive attitudes emerged, as 'greater knowledge about the subject and exposure to the ideas tends to be related to acceptance' (Ipsos MORI, 2016). This suggests the possibility of shifting public perceptions and consequently of data provider attitudes to risk.

Conclusion

This study evidences the variations in practice across government in relation to consent for the secondary research use of administrative data, articulated around six factors. In unpacking these variations, there is a fundamental and unanswered question, which is whether consent is in fact the best mechanism for enabling individual privacy and public protection against harm. Manson and O'Neill (2007) argue in the context of biomedical ethics that consent has become an accepted orthodoxy. Consent mechanisms are the primary means of ensuring the individual's right to choose (Laurie and Postan, 2013). When individualistic self-determination is the goal of data governance frameworks, the individual gains highly specified degrees of choice over how 'their' data is used. This is echoed in our study:

> …risk is highly contextual, and actually individuals are reasonably good at understanding the benefits and rewards of different things. We make trade off decisions

> all the time, and we do so really remarkably well on the whole, but the professional attitude is that you aren't in a position to calculate those risks ...I am much more trusting of the public in their abilities to make decisions. (A36)

Autonomy through highly specific control by the data subject enables dynamic consent in settings such as biobanks (Budin-Ljøsne et al., 2017; Kaye et al., 2015). In an energy context, the introduction of smart meters gives a greater degree of control over data sharing to the homeowner. However, a highly individualised approach to data sharing, based on the granular consent of the data subject, is not the only model for government administrative data re-use for research.

The majority of researchers we interviewed placed individual consent as one among many governance mechanisms within a social contract model of data sharing. In this framing, the end goal is protection of data subjects from deception and harm. Interviewees who championed a social contract model of data sharing framed personal data, not through the language of individual ownership, but of co-production:

> People have an image of medical data and confidentiality that is often captured by the language of ownership. I have it, I give it to the GP to use, and if they don't use it in the way I expect then that is an abuse. But actually I don't have it, the data is part of the relationship, all this data is co-produced... We need to understand what the ground rules are of that co-production process, and we need to work out what is a betrayal of those ground rules and what isn't, and that requires a lot more transparency and openness but it doesn't necessarily require a high degree of individual privacy choices. (A42)

In archival science, Iacovino (2010) and others have considered a participant relationship model which 'acknowledges all parties to a transaction as immediate parties with negotiated rights and responsibilities'. Such a social model of co-production of data re-conceptualises the data subject as a data co-creator and acknowledges both individual and institutional (indeed, multiple stakeholder) rights in the data.

As phrased by A42, rather than focusing on consent, 'it would be much better to think about permissions and licences and people's agreement to trust certain gateway organisations'. In this model:

> We need to build up an account of what makes it acceptable to use this information through a combination of things, which is not just about 'we want to do this, do you agree?' but it is more about 'we are not sure what we might like to do with it, we would like you to be part of the enterprise, this is what we are building in as safeguards and protections'. ...We should be thinking about a 'no surprises' rule, in that, even if people haven't set out expecting this, if we have reason to believe that they would be surprised and upset then we need to put some effort into being open and transparent about it. But I don't think it is helpful to think about that as a question of consent, it is more a question of governance and acceptance. ...We need ... a social contract that allows us to say that we feel we have permission to do these things... we have to put in place an open and transparent explanation of what we think we are doing, some processes to reassure that we are doing what we thought and said we are doing, and some mechanisms for accountability. (A42)

Governance mechanisms for data sharing must ultimately work to achieve a balance between the mutually reinforcing public goods of protecting privacy *and* enabling use that is in the public interest. Yet, a growing body of academics, in the context of Big Data, are questioning whether consent as a governance mechanism can enable this balancing act (Barocas and Nissenbaum, 2014a, 2014b; Cate and Mayer-Schönberger, 2013; Kuner et al., 2012; Rubinstein, 2013). Even in isolating the issue of privacy, the potential in Big Data challenges an individual-orientated governance approach. Barocas and Nissenbaum explain this through an exploration of the 'tyranny of the minority' (2014a), where consented disclosure of information by a few can reveal information about the many. Barocas and Nissenbaum explore consent within its social landscape and argue that in both academic and regulatory circles, attention has focused on seeking to improve procedures and mechanisms for capturing informed consent (2014a). Drawing on Manson and O'Neill (2007), Barocas and Nissenbaum (2014a) argue instead for a greater focus on mechanisms that propagate the social acceptability of data re-use through a shared articulation of the underpinning rights, obligations, and legitimate expectations surrounding re-using data.

In practical terms, what might a model based on a shared articulation of this kind look like? The Nuffield Council on Bioethics (2015) proposed a social contract model of data sharing. A social contract model would need to be underpinned by an agreed set of reasonable expectations about how data will be shared; clarity and transparency about the process by which individual freedoms are respected; agreement on the governance that will give acceptable assurances, and on who is accountable for what.

The ethical issues concern the privacy of individuals and the risk of disclosure, but also the larger moral consequences and social impact. The balance of risks

must ensure that data is used responsibly to promote the public interest, in a way that best reconciles the interests of individuals and groups, in keeping with their fundamental rights (Nuffield Council on Bioethics, 2015). Consent is designed to convince data subjects and public stakeholders of a pre-determined public good in research. In contrast, co-constructed participation in a deliberative and dynamic process balances the relationship between public and private interests: this might move the debate towards Rawls's (1958) question of justification. A social contract model 'recognises the necessarily provisional nature of decisions about data management and governance, since the horizon of possibilities – and the values and interests invested in them – are constantly changing as the social, political, technological and information environments evolve' (Nuffield Council on Bioethics, 2015).

In conclusion, then, our study suggests that there is no single agreed formula for the use of consent as a sufficient mechanism to ensure privacy in the secondary use of government administrative data. In seeking to identify applicable norms, mere compliance with the law and adoption of one-time consent processes is inadequate to ensure that data use is ethical and morally reasonable. Information governance frameworks must go beyond the law, based on an identification of reasonable expectations of privacy and data use held by all interested parties, to determine the social thresholds for what is acceptable in a given use context. Consideration of the six factors identified in this research could lead to the development of a more reflexive and dynamic process of articulating the justification for data sharing and re-use, moving towards a social contract model of agreement which would provide a more trusted and transparent approach to sharing of government administrative data. This would go some way towards addressing the limitations which our research, and the work of others, has highlighted of the use of consent as the primary governance mechanism. Our research suggests that consent is no longer the best or only mechanism for data re-use and suggests that a dynamic social contract model might provide a better approach.

Acknowledgements

The authors are grateful to members of the sub-project Management Group (ADRCE-project 003.01) for their advice and guidance: Professor Ruth Gilbert (UCL Institute of Child Health), Professor Lorraine Dearden (UCL and Institute for Fiscal Studies) and Professor Paul Wilkinson (The London School of Hygiene & Tropical Medicine).

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This work was funded by the UK Economic and Social Research Council (ESRC) as part of the ADRC-E (Administrative Data Research Centre – England) project, 2013–2018, grant number ES/L007517/1.

ORCID iD

Elizabeth Shepherd http://orcid.org/0000-0003-2404-0149

References

Barocas S and Nissenbaum H (2014a) Big data's end run around anonymity and consent. In: Lane J, Stodden V and Bender S (eds) *Privacy, Big Data, and the Public Good: Frameworks for Engagement*. New York: Cambridge University Press, pp. 44–75.

Barocas S and Nissenbaum H (2014b) Computing ethics: Big data's end run around procedural privacy protections. *Association for Computing Machinery. Communications of the ACM* 57(11): 31–33.

BBC News (2012) DVLA bans councils from database over abuses. *BBC News*. Available at: www.bbc.co.uk/news/uk-politics-20642429 (accessed 7 December 2018).

Budin-Ljøsne I, Teare HJ, Kaye J, et al. (2017) Dynamic consent: A potential solution to some of the challenges of modern biomedical research. *BMC Medical Ethics* 18(4) DOI 10.1186/s12910-016-0162-9. Available at: https://bmcmedethics.biomedcentral.com/articles/10.1186/s12910-016-0162-9 (accessed 7 December 2018).

Caldicott F (1997) *Report on the Review of Patient Identifiable Information*. London: Department of Health.

Caldicott F (2016) Review of data security, consent and opt-outs. *National Data Guardian*. Available at: www.gov.uk/government/publications/review-of-data-security-consent-and-opt-outs (accessed 7 December 2018).

Carter P, Laurie GT and Dixon-Woods M (2015) The social license for research: Why care.data ran into trouble. *Journal of Medical Ethics* 41: 404–409.

Cate F and Mayer-Schönberger V (2013) Notice and consent in a world of big data. *International Data Privacy Law* 3(2): 67–73.

Critchley R, Gilbertson J, Grimsley M, et al. (2007) Living in cold homes after heating improvements: Evidence from warm-front, England's home energy efficiency scheme. *Applied Energy* 84(2): 147–158.

Flinn A and Shepherd E (2011) Special issue: Archives, records and identities – Questions of trust. *Archival Science* 11: 3–4.

Foulds C and Powell J (2014) Using the homes energy efficiency database as a research resource for residential insulation improvements. *Energy Policy* 69: 57–72.

General Data Protection Regulation (GDPR) (2016) *Official Journal of the European Union*. Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN (accessed 7 December 2018).

Grace J and Taylor MJ (2013) Disclosure of confidential patient information and the duty to consult: The role of the health and social care information centre. *Medical Law Review* 21(3): 415–447.

Health e-Research Centre (2016) Citizens' jury: Health data on trial. Available at: www.herc.ac.uk/get-involved/citizens-jury/ (accessed 7 December 2018).

Health & Social Care Information Centre (HSCIC) (2015) Supplementary written evidence Parliamentary Health Committee. Available at: http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/health-committee/handling-of-nhs-patient-data/written/17671.html (accessed 7 December 2018).

Health & Social Care Information Centre (HSCIC) (2016) Applying type 2 opt-outs. Available at: https://webarchive.nationalarchives.gov.uk/20180307201304/http://content.digital.nhs.uk/article/7072/Applying-Type-2-Opt-Outs (accessed 7 December 2018).

Iacovino L (2010) Rethinking archival, ethical and legal frameworks for records of Indigenous Australian communities: A participant relationship model of rights and responsibilities. *Archival Science* 10: 353–372.

infiniteideasmachine (2012) Blogpost, *Opening the NPD*. Available at: www.infiniteideasmachine.com/2012/12/opening-the-national-pupil-database-a-parents-response/ (accessed 7 December 2018).

Information Commissioner's Office (2017) Consultation: GDPR consent guidance. Available at: https://ico.org.uk/media/about-the-ico/consultations/2013551/draft-gdpr-consent-guidance-for-consultation-201703.pdf (accessed 7 December 2018).

Ipsos MORI (2014) Dialogue on data: Exploring the public's views on using administrative data for research purposes. Available at: https://www.ipsos.com/ipsos-mori/en-uk/dialogue-data (accessed 7 December 2018).

Ipsos MORI (2016) The One-Way Mirror: Public attitudes to commercial access to health data. Available at: https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf (accessed 7 December 2018).

Kaye J, Whitley EA, Lund D, et al. (2015) Dynamic consent: A patient interface for twenty-first century research networks. *European Journal of Human Genetics* 23(2): 141–146.

Kuner C, Cate F, Millard C, et al. (2012) The challenge of 'Big Data' for data protection. *International Data Privacy Law* 2(2): 47–49.

Landshoff P and Polak J (2008) The National Transport Data Framework. Available at: https://www.repository.cam.ac.uk/handle/1810/198269 (accessed 7 December 2018).

Laurie G and Postan E (2013) Rhetoric or reality: What is the legal status of the consent form in health-related research? *Medical Law Review* 21(3): 371–414.

McDermott Y (2017) Conceptualising the Right to data protection in an era of big data. *Big Data & Society* 4(1) DOI: 10.1177/2053951716686994. Available at: https://journals.sagepub.com/doi/full/10.1177/2053951716686994 (accessed 7 December 2018).

MacNeil H (2011) Trust and professional identity: Narratives, counter-narratives and lingering ambiguities. *Archival Science* 11(3–4): 175–192.

Manson NC and O'Neill O (2007) *Rethinking Informed Consent in Bioethics*. Cambridge: Cambridge University Press.

Nuffield Council on Bioethics (2015) The collection, linking and use of data in biomedical research and health care: Ethical issues. Available at: https://nuffieldbioethics.org/wp-content/uploads/Biological_and_health_data_web.pdf (accessed 7 December 2018).

NHS Constitution for England (2015) Available at: www.gov.uk/government/publications/the-nhs-constitution-for-england/the-nhs-constitution-for-england (accessed 7 December 2018).

O'Neill O (2003) Some limits of informed consent. *Journal of Medical Ethics* 29: 4–7.

Ostherr K, Borodina S, Conrad Bracken R, et al. (2017) Trust and privacy in the context of user-generated health data. *Big Data & Society* January–June 2017: 1–11. DOI: 10.1177/2053951717704673. Available at: https://journals.sagepub.com/doi/10.1177/2053951717704673 (accessed 7 December 2018).

Partridge N (2014) *Review of Data Releases by the NHS Information Centre*. Health and Social Care Information Centre. Available at: https://www.gov.uk/government/publications/review-of-data-releases-made-by-the-nhs-information-centre (accessed 7 December 2018).

Rawls J (1958) Justice as fairness. *Philosophical Review* 67(2): 164–194.

Rubinstein I (2013) Big data: The end of privacy or a new beginning? *International Data Privacy Law* 3(2): 74–87.

Ruppert E, Law J and Savage M (2013) Reassembling social science methods: The challenge of digital devices. *Theory, Culture & Society* 30(4): 22–46.

Sexton A, Shepherd E, Duke-Williams O, et al. (2017) A balance of trust in the use of government administrative data. *Archival Science* 17(4): 305–330.

Shakespeare S (2013) An independent review of public sector information. Available at: www.gov.uk/government/publications/shakespeare-review-of-public-sector-information (accessed 7 December 2018).

Stanford Encyclopedia of Philosophy (2017) Contemporary approaches to the social contract. Available at: https://plato.stanford.edu/entries/contractarianism-contemporary/ (accessed 7 December 2018).

Taylor MJ (2015) Legal bases for disclosing confidential patient information for public health: Distinguishing between health protection and health improvement. *Medical Law Review* 23(3): 348–374.

Taylor MJ and Taylor N (2014) Health research access to personal confidential data in England and Wales: Assessing any gap in public attitude between preferable and acceptable models of consent. *Life Sciences, Society and Policy* 10(1): DOI: 10.1186/s40504-014-0015-6. Available at: https://lsspjournal.biomedcentral.com/articles/10.1186/s40504-014-0015-6 (accessed 7 December 2018).

UK Department for Business, Energy & Industrial Strategy (BEIS) (2013) National Energy Efficiency

Data-Framework (NEED). Available at: www.gov.uk/government/collections/national-energy-efficiency-data-need-framework (accessed 7 December 2018).

UK Department for Education (DfE) (2016) Data management advisory panel (DMAP): Terms of reference. Available at: http://dera.ioe.ac.uk/id/eprint/26863 (accessed 7 December 2018).

UK Department for Transport (DfT) (2013) Open data strategy refresh. Available at: https://www.gov.uk/government/publications/open-data-strategy-refresh (accessed 7 December 2018).

UK Government (2002) The road vehicles (Registration and Licensing) regulations 2002 SI 2002 no. 2742. Available at: www.legislation.gov.uk/uksi/2002/2742/contents/made (accessed 7 December 2018).

UK Government (2013) The Education (Individual Pupil Information (Prescribed Persons) (England) Regulations 2009 SI 2009 no. 1563, as amended by SI 2013 no. 1193. Available at: http://legislation.data.gov.uk/cy/uksi/2013/1193/made/data.htm?wrap = true (accessed 7 December 2018).

Vezyridis P and Timmons S (2017) Understanding the care.data conundrum: New information flows for economic growth. *Big Data & Society* 4(1) DOI: 10.1177/2053951716688490. Available at: https://journals.sagepub.com/doi/abs/10.1177/2053951716688490 (accessed 7 December 2018).