



28 Bedford Square
London WC1B 3JS
Telephone: +44 (0)20 7631 0566
Fax: +44 (0)20 7323 4877
www.nuffieldfoundation.org
Registered Charity 206601

Better Use of Data in Government Consultation 2016
Government Digital Service
6th Floor
Aviation House
London
WC2B 6NH

By email: to data-sharing@cabinetoffice.gov.uk

22 April 2016

Dear Sir/Madam

This letter sets out the Nuffield Foundation's response to the consultation on better use of data in government.

The Nuffield Foundation is a charitable trust established by William Morris, Lord Nuffield, the founder of Morris Motors. Our aim is to improve social well-being and we do this by:

- Funding research and innovation projects in education and social policy, with a strong emphasis on quantitative analysis using existing (and government-managed) datasets. In 2015 we funded 34 new projects with a total value of £5 million. Seventy-five per cent of these projects were based in universities.
- Building research capacity in science and social science, most notably through Q-Step, a £19.5m programme designed to promote a step-change in quantitative social science training for undergraduates (co-funded with the ESRC and the Higher Education Funding Council for England (HEFCE)), and Nuffield Research Placements, which provide over 1,000 Year 12 students each year with the opportunity to undertake STEM research projects.

Relevance to our work

Many of the research projects that we fund draw upon administrative data, and researchers are often looking for novel ways to use multiple sets of data to inform policy. As such, access to, and analysis of, data which is gathered by government departments, arm's length bodies and allied third parties, is essential to our work. We also recognise that in many of the policy areas in which data can shed light or (at least) help to design future studies, there is no single source of data, and therefore there is great potential for aligning and joining datasets. The recent Nuffield-funded research on the 'human capital' of graduates is a pertinent

example of this¹. For the first time, researchers were able to use anonymised tax data and student loan records to look at how graduate earnings vary by institution of study, degree subject and parental income up to ten years after graduation. Of course, we are aware that some data are more sensitive than others and that different protocols might affect how these data are treated compared with, say, 'open data'.

Some areas of the consultation were more relevant to our work than others and for that reason we are responding to a sub-set of questions rather than the entire suite. The remainder of this letter sets out the views of the Nuffield Foundation on these substantive areas, offering views on the general (consultation questions 1-4) and the specific (consultation questions 15-20) areas.

General points (consultation questions 1-4)

It is our belief that issues relating to data access and linking should start from a presumption of permissiveness and that whichever authority is brought into being to manage access should adopt that position. Data requests should, of course, make a convincing case for the data required, but there should be clarity over the evaluation criteria and an onus on the relevant body to make a comprehensible and visible case for refusing requests.

We welcome the safeguards put in place to ensure that Ministers must consult widely before adding new objectives (although a-c on page 11 of the consultation seem so broad in scope as to make this power redundant). Ensuring this is not subject to 'whim' is an important step. We support the availability of data for research purposes, and it is also our view that under this legislation, the position of non-public sector bodies undertaking public services (for example Free Schools and Academies) needs to be consistent with that on public sector bodies, subject to controls over the security of data used by such organisations.

There appears to be no reference to the extent to which longitudinal and survey work can be included in the scope of this legislation and we would suggest that it should be. Linking such exercises to administrative data, for example, could be very powerful, and would have a number of benefits. First, use of administrative data for the purposes of sampling (as opposed to analysis) can be an efficient and cost-effective way of enabling researchers to identify populations of special interest who may be missing from general household surveys. This has the dual advantage of making better use of limited research budgets (which will be doubly important if new charges are proposed for data access) and of enabling the participation of hard-to-reach groups (whose under-representation in surveys is not always due to different response modes – some groups are missed because they are hard to find in the first place). Second, where survey and administrative data can be linked, this can improve data quality through triangulation of self-reported responses with administrative data, and ease respondent burden through reducing the number of questions asked, or the number of times a person is surveyed (if they can subsequently be monitored through administrative data).

We appreciate that these uses of administrative data for survey purposes raise slightly different challenges, because the researcher may need to access personal – as well as anonymised – data, but we think it is important that they receive greater attention. To that

¹ <http://www.ifs.org.uk/publications/8233>

end the Foundation, in partnership with the Economic and Social Research Council (ESRC) and the Medical Research Council (MRC)-funded CLOSER unit at University College London (UCL), is holding a seminar on 13 July to explore what might be done to make progress on these issues.

Equally, the consultation needs to be closely aligned with the protocols and considerations over the security of individuals' data which are emerging from the Caldicott review.² It is likely that principles which are fit for the sharing of health and social care will be transferable to other forms of personal data.

We also take the view that the legislation needs to avoid tying itself closely to forms of technology (for example, Application Programming Interfaces (API)) and specific mechanisms for linking data (such as described in paragraph 98). This is because the technology is moving so fast that there are risks that it will not stand the test of time. The legislation needs to focus upon comprehensible and forward-looking principles, with examples of what to use and how to do things made available in accompanying guidance (which can be updated as necessary).

There is no timetable proposed for implementation of the legislation and whilst that may be indicative of caution, it would be helpful to have some outline so that the implications can be planned for. It is also our view that the consultation document conveys the impression that data and intelligence collection are a burden and that the benefits are solely economic or efficiency inclined. As with the recent Office for National Statistics (ONS) consultation 'on Changes to ONS Products 2015', we feel that the broader social benefits (the 'public good') of intelligent data gathering, linking and sharing (with appropriate safeguards) need to be given stronger weighting.

Specific points (consultation questions 15-20)

Many of the above points have a direct bearing upon the ways in which researchers would like to use data gathered by government (for example, the legislation should include linking to survey data). However, the questions raised in relation to research require some additional scrutiny.

The creation of a body to oversee 'trusted third parties' is an attractive concept, especially if it is also able to broker and manage sharing arrangements (with a presumption of accessibility). There are, however, some logistical and financial considerations which may influence the effectiveness of such a body. Some reassurance - perhaps drawing on the experience of the Administrative Data Service (ADS) - about the current volume of data linking requests which are made to government departments and the speed of turnaround (and the costs involved) would be very helpful here.

Related to this, our main concerns would be the speed with which the requests could be processed, and the extent to which the new body can support ('compel') the development of new links. The majority of our research grants cover a period of 18-24 months, so processing requests swiftly and reliably is important. Although we have numerous positive

2

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/461993/Consultation_ND_Guardian_accessible.pdf

examples, we and our grant-holders also have considerable experience of slow and sometimes seemingly obstructive responses to data access requests from responsible bodies. This creates some concern about the extent to which a new layer of data administration will be able to act to ensure that public benefit (however that is defined, but preferably on a case-by-case basis) data requests are managed in a timely fashion.

More generally we have concerns about the creation of a system of approvals from which government departments and agencies across the UK themselves may be exempt (when they are wishing to access and analyse data). We regularly hear of instances where government departments have not adhered to rules under which they expect external researchers to operate. Whether or not these examples are anecdotal or have substance, the fact that they are in circulation does nothing to build a culture of consistency and mutual trust, which we would argue are essential foundations for responsible use of data.

We are moving forward with a proposal to develop a Family Justice Observatory,³ and this will draw heavily on data and surveys from similar sources as set out in the consultation concerning the Troubled Families programme. We are hopeful that the proposed legislation will be able to support this new enterprise in its mission to provide access to reliable research to inform analysis of options for specific cases and system performance data to assist in the allocation of resources and the development of new services and interventions. There is certainly strong demand across the system – which extends well beyond government departments – for improved understanding of both the system itself and of the drivers of demand. The Nuffield-funded work of Professor Karen Broadhurst at Lancaster University, who was able to use data from the Children & Family Court Advice and Support Service (Cafcass) to explore the extent to which mothers subject to care proceedings are recycled through the system, has brought long-overdue focus on the need to address the underlying problems that generate a high proportion of cases.

These matters link back to the resources available to any new body, and we note the suggestion about charging a fee for access and processing. Again, there are implications for a grant-making organisation such as the Nuffield Foundation as this fee will have to be included as part of any grant proposals. So clarity over the pricing structure needs to be available (and tested) early so that applicants can plan for these additional costs when making applications for funding. In principle, the Foundation is not wholly opposed to making a reasonable contribution to improved data access, but we are not convinced that doing this on a ‘pay as you go’ basis is the right approach, not least as the administrative burden for the data-owners is unlikely to address the need for a more permissive approach. More importantly – given that administrative data are often incomplete or of variable quality – we are concerned that a ‘pay-as-you-go’ model may in fact create disincentives for departments to address underlying structural/quality issues with the data if they can simply charge for the work needed to undertake temporary fixes. In addition, it will be difficult to plan for income received from a ‘pay-as-you-go’ model and this will make it even harder to plan for improvements to infrastructure. We also recognise that much of the effort and cost in creating new linked sets of data arises the first time it is generated. De-identified linked datasets should, as a matter of course, be made available to a wider audience to offset the costs of re-creating them.

³ <http://www.nuffieldfoundation.org/news/team-lancaster-university-and-alliance-useful-evidence-establish-scope-and-delivery-new-family->

If fees are to be set, however, it would be advisable to make this scalable and to set limits based on either modelling scenarios or transferable experience (for example from Scotland). Data and linked data are obviously not a 'no cost' activity for government and some recognition of this is probably inevitable, but the introduction of fees is likely to have the effect of raising expectations of an improved service. We would therefore expect that any income generated should be directed back into service improvement and (for example) supporting the creation of generalised and regularly updated de-identified data sets (for example, the National Pupil Database – NPD).

A consequence of creating any 'accrediting' body is that it will have to police compliance with whatever code of practice it establishes and also be capable of removing recognition as an approved indexer, researcher or other data user. It is also worth considering whether this recognition could be given at organisation-level with a concomitant responsibility on that organisation to evaluate compliance with due procedures. It is not clear how frequently such re-accreditation would have to take place, but we would suggest that it should be at the very most an annual process or one which is proportionate to the track record of the organisation and/or individuals.

Lastly, the consultation asks about the publication of rejected applications. The availability of such information (without identifying the source of the request) will prove to be valuable both in terms of establishing the reliability of the application of any agreed access principles and setting out clearly what is and is not practical. Over time, this may become a reference for a working definition of 'public interest'.

We hope you will find these views useful and would be very happy to provide more details on any of the issues discussed in this response.

Yours sincerely,

A handwritten signature in black ink that reads "Josh Hillman". The signature is written in a cursive, slightly slanted style.

Josh Hillman
Acting Director