

Academy of Social Sciences response to Cabinet Office consultation on *Better Use of Data in government*

1. The Academy of Social Sciences is the national academy of academics, learned societies and practitioners in the social sciences. Its mission is to promote social science in the United Kingdom for public benefit. It is composed of around 1,000 individual Fellows, 41 Learned Societies, and a number of affiliate members, together representing nearly 90,000 social scientists.

Introduction

2. The Academy's focus in this response is on Section C of the Cabinet Office Consultation on *Better Use of Data*, on 'allowing use of data for research and for official statistics'. This section relates to 'de-identified' data for research and identifiable data for national statistical purposes. The Academy believes that the privacy implications in these two areas are rather different from the sharing of personally identifiable data between government departments for the other purposes set out in the consultation. At earlier stages in the open policy-making process, the Academy raised the issue of whether legislation about de-identified data for research and data for statistical purposes should be considered separately from legislation containing provisions for sharing personally identifiable data between government departments. We strongly believe this issue bears further consideration to ensure appropriate parliamentary consideration of the issues involved, as the issues in each area are complex though different from each other.
3. The Academy welcomes the prospect of legislation that may help promote clarity about the use of administrative data for research and statistical purposes. But, as we discuss below, we are not convinced that legislation alone will promote the cultural change that will, we believe, be necessary in order to reap the benefits of wider access to data for research and statistical purposes. Access to government data for research, even under existing mechanisms such as the Administrative Data Research Network, which has served as a model for some parts of the proposals contained in the consultation, has been frustratingly slow, often for very opaque reasons. We return to this issue in paragraph 10 below, in our discussion of incentives for access to data.
4. UK social science has long had a robust and transparent regime for data linkages. This involves participant consent where social surveys are being linked to administrative data, and methods of ensuring public scrutiny for linkages of administrative data under the Administrative Data Research Network (ADRN), set up by the Economic and Social Research Council and governed by a Board convened by the UK Statistics Authority. It is important to note from the outset that the ADRN has robust mechanisms for considering not only the scientific appropriateness of proposals for data linkage but also the ethical and privacy implications of any proposal, and that these involve the participation of lay members (not just social scientists or statisticians). The ADRN publicly and transparently lists

all approved projects to help enable public accountability and scrutiny. And the protections put into place involve provisions other than the nature of the de-identified data to help safeguard citizen privacy. These may be helpfully characterised as the 'five safes', following the Office for National Statistics and the ESRC-funded UK Data Service, <http://blog.ukdataservice.ac.uk/access-to-sensitive-data-for-research-the-5-safes/>. Provisions involve:

- Safe projects: for public good, as judged by a transparent mechanism independent of the researchers, and including lay members;
 - Safe data: de-identified (stripped of personal identifiers such as name and address and other data that pose a disproportionate risk of personal identification);
 - Safe people: through accredited and approved researcher schemes, involving training, clear and enforceable rules, and sanctions both of institutions and researchers who do not follow them;
 - Safe settings: secure settings for holding and processing data, following agreed international security safety standards;
 - Safe outputs: publication of results that have been specially screened to ensure that they cannot inadvertently provide information which could reveal personal identities.
5. The Academy is pleased that many of the safeguards for research and statistical use of administrative and other data proposed in the consultation are based on the models provided by UK social science, largely supported by the Economic and Social Research Council. We welcome too the clear statement of the 'key protective principles' set out in (the second) paragraph 13, on page 5 of the consultation document.

Access by the UK Statistics Authority to data for the purpose of producing official statistics and research

6. The Academy of Social Sciences supports the work of the Royal Statistical Society in promoting its Data Manifesto, which sets out the importance of robust statistics in democratic societies. We concur with them that the current framing of the draft legislation regarding access by the Office of National Statistics to government departmental data is not sufficient to ensure the access required for robust national statistics. This issue has also been brought to prominence by Sir Charles Bean's *Independent Review of UK Economic Statistics*. We agree with the recommendation in the Bean Review that ONS should have a similar right of access to that accorded to the Office of Budget Responsibility, with an appropriate form of independent oversight in contested cases. We note too that stronger data sharing rights are not only important for the production of timely and accurate national statistics, but are essential to achieve the reduction in respondent burden and the financial savings envisaged in the consultation document.

Access to de-identified data for research purposes

7. The Academy understands why access to de-identified data for research purposes should be based on a permissive power given to government departments and data owners. But we have concerns that the current reluctance on the part of many departments to provide data for research that has been through approved mechanisms such as the ADRN (about which the ESRC could provide details) may persist. If this legislation is to promote real access, and not just rhetoric, will, we believe require incentives for cultural change.
8. This is despite public views (see the report commissioned by the ESRC and ONS on the issue of using administrative data for research, <https://www.ipsos-mori.com/Assets/Docs/Publications/sri-dialogue-on-data-2014.pdf>) supporting the use of administrative and other data for research purposes as long as appropriate safeguards are in place. The Academy understands the complexity of making data available for social scientists' use that may address issues of contemporary social policy. It supports the use of independent and peer review mechanisms to ensure that proposed projects using government data (and indeed all data) are not only scientifically robust but provide objective analyses of complex issues. But we are concerned that governments of any complexion are often concerned that wider use of data may bring government policies under scrutiny, and the legislation could provide important incentives to promote cultural change.
9. This issue may be illustrated by the proposed Code of Practice for statistics and research, on page 18, in which it discusses the additional conditions that the UKSA are likely to set out for research, including that any proposals 'not be research which a government department or agency would carry out as part of its normal operations'. The Academy is concerned that this provision is too widely drawn. As currently drafted, government departments could justify decisions not to allow data access if the project were on an issue that was of any policy relevance. Given that data will always be provided in arrears, after a lengthy proposal process, and strict scrutiny of its objectivity, we think a redrafting is needed to ensure that data requests are not denied because the relevant department might normally carry out such work, or might carry out such work but might not publish it. The public benefit test should be sufficient. We believe that this clause should be re-drafted to focus on 'wholly managerial or administrative operational matters', to ensure that policy research was not included.
10. Other proposals in the Consultation Paper may help address the issue of incentives for permitting access. The first relates to listing proposals for data access. **Question 16** of the consultation asks whether the UK Statistics Authority, as the accrediting authority, should publish details of rejected applications for data access, along with the reasons for their rejection. We would not favour listing full details (for instance, including researcher names) of those applications rejected by the independent adjudication mechanism as used by ADRN, which have fallen at the hurdles of scientific rigour or objectivity or public benefit. There could however be

a full list with details of all proposals approved by an independent body (like the mechanisms set up by the ADRN), including a brief description of the proposal, showing which department's data was requested, and whether the final departmental decision was to allow access or not. It would also be helpful to have information about the date at which initial approval by the independent panel was granted, and the date when data was provided by the relevant government department or the decision was made not to do so, including the reasons given if access for an approved project was not granted. Such a report, if compiled on an annual basis, with oversight and publication by UKSA and perhaps put forward as part of its annual parliamentary report, would ensure that there was continuing scrutiny of the departmental practices and their reasoning, and that government departments faced some accountability mechanism in making their decisions about whether or not to grant access to requests that have been approved by the independent mechanism as meeting the public good test (and being scientifically robust).

11. In any event, we stress the importance of resourcing the UKSA appropriately for this new role in supporting the adjudication mechanism for this work.

Details of proposed data access mechanisms

12. We welcome the careful consideration of the mechanisms by which data requests might be made, the sort of scrutiny they will need, and the mechanisms for providing data contained in the Consultation Paper and the Impact Statement on de-identified data for research. We have some concerns that these are perhaps too specifically drawn in some cases (for instance in paragraphs 97 and 98, referring to Trusted Third Parties, the current mechanism for de-identification and data linkage). We wonder if the drafting should be reconsidered, setting out a standard at least as strong as the Trusted Third Party model, leaving details to be set out in secondary legislation (as proposed) but not putting detail which could become obsolete into the proposed Bill.
13. The Academy is particularly concerned that destruction of *de-identified* data not be mandated in legislation but left to secondary legislation. Given the administrative resources required for each access, and the fact that research uses may not always require the most recent data, we can see possibilities for re-use of de-identified data (after going through the process for independent scrutiny of each proposal for re-use and subject to the relevant security provisions for de-identified data).
14. This holds in particular if the proposed legislation brings the linkage of survey data and administrative data within its scope, as we would hope. To date, some of the richest uses of departmental data have arisen from the links between UK's unrivalled collection of longitudinal survey data (such as those run from the ISER at the University of Essex and the birth cohort studies overseen by CLOSER, at University College London – Institute of Education) and administrative data. Once the administrative data are de-identified, and are linked to the survey data, they are held

in secure settings and adhere to the other five 'safes' listed on paragraph 4 above. We do not believe that this contravenes the key principles set out by the Consultation Paper.

15. The Academy notes that the possible inclusion of health and social care data (including data held by the Health and Social Care Information Centre) is awaiting the findings of Dame Fiona Caldicott's review, but we think it important that if there are to be two regimes they should permit the linkage of administrative data for social research with health data, so that unintended obstacles are not created by the legislation. We believe that linking social administrative and survey data with health data will become increasingly important in addressing many issues in preventative health care and behavioural change, an issue of potentially enormous public benefit.
16. **Question 15** in the Consultation Paper raises the issue of fees that might be charged by public authorities to provide data for research purposes. The Academy understands why cost-recovery may be an issue. Ideally we would urge that the UKSA be given funds to support this work for approved projects, at least for the first few years. In principle we would make the point that funding research for public good is an appropriate role for government, and we suspect the total sums involved would not be large. But if fees are to be charged, there needs to be some work, with appropriate oversight, to establish a clear and appropriate (and we think in reality modest) tariff for provision of data. We think it important that any tariff reflect only the marginal costs of the data preparation work and that there is some scrutiny and cross-departmental consistency in this. Otherwise, departments could set inappropriately high charges to deter requests, including asking researchers to bear the costs of protracted departmental deliberations. Since most of the research requests that might make use of data for public benefit would be funded by research councils (funded by government) or research charities who fund such work as a public good, the Academy believes it is in the public interest, if costs are not to be subsidised by government as a public good, that they are kept to a minimum, and a clear tariff is set in advance that allows research funders to plan accordingly.

Definition of public benefit

17. Regarding **Question 17** of the Consultation, the Academy agrees that the definition of 'public good' as set out in the Statistics and Registration Act 2007, resting on 'informing the public about social and economic matters' and 'assisting in the development and evaluation of public policy', is a sound substantive approach, and encompasses a wide range of possible uses.
18. The Academy has long been a proponent of the view that the concept of 'public benefit' should be drawn to include certain procedural safeguards. These include that decisions about the public benefit test for permitting access should be defined by adherence to a process including consideration of the proposal independently of the researcher or funder, and involve lay members of the public (that is, not social

scientists or government staff), who must decide if there is a legitimate public benefit, and balance the benefits and any potential harms. Further, we believe that an important safeguard is to publish the decisions about both recommendations that access be given, and subsequent departmental decisions either to permit or deny access, as set out in paragraph 10 above. This allows public scrutiny and a mechanism for accountability. A final test would be that research findings should be published, so placing them in the public domain.

19. The Academy believes that these tests might allow further discussion of allowing access to 'commercial users' doing public benefit research. We note the recent publication of a report by the Wellcome Foundation, to inform the deliberations of the Caldicott review, again carried out by Ipsos MORI, on public attitudes to commercial access to health data. (See <https://www.ipsos-mori.com/researchpublications/publications/1803/Commercial-access-to-health-data.aspx>.)
20. That report may help in due course to distinguish between research users and research uses, and raises many important issues. We believe that adjudication by an independent body with lay members, open listing of permissions given and requests granted, and publication of research results are a starting point in judging whether there is a genuine public benefit, even in the case of commercial users.
21. But public views on this matter have been affected by the reaction to Care.data. It will be important for any legislation to be clear about whether commercial users are or are not included; their inclusion at this stage may require further time and discussion of the Caldicott review. If this were to lead to a protracted delay in putting forward legislation, it would be better to restrict this legislation to approved users as set out by the ESRC and other research councils.