

Cabinet Office consultation on 'Better use of data in government'

Response from the CLOSER Consortium

CLOSER (Cohorts and Longitudinal Studies Enhancement Resources) is funded by the Economic and Social Research Council (ESRC) and Medical Research Council (MRC) to maximise the use, value and impact of the UK's longitudinal studies. The network brings together eight leading longitudinal studies, the British Library and the UK Data Archive.

CLOSER's response to the consultation predominantly reflects our interest in being able to link individual administrative records to longitudinal survey data collected from that same individual. The linking of these two forms of data offers potential that exceeds the sum of the two approaches in isolation. While administrative records can provide enviable coverage and frequency, they rarely provide the richness and depth that comes from survey data collected to tackle specific research questions; conversely, survey data collection can be costly and challenging to administer, leading to missing data and possible biases which can be overcome by using administrative records to fill in the missing information. Consequently, combining the two can yield considerable benefits.

For example, recent analysis of the National Pupil Database (NPD) and Higher Education Statistics Agency (HESA) found higher participation rates in higher education among a range of groups including ethnic-minorities.¹ However, those records could not help explain these differences in participation; exploring this required linking individual administrative records to survey data collected as part of the Avon Longitudinal Study of Parents and Children and the Longitudinal Survey of Young People in England. This linkage yielded important insights about the interplay between gender, ethnicity and socioeconomic status and wider social, cultural, personal and economic factors.² It is notable that the recent House of Lords Select Committee report identifies exploring this very issue – patterns of entry to higher education among young people – as an area where “existing data is unreliable and inconsistent” and strongly supports further sharing of administrative data.³

¹ Crawford C and Greaves E. (2015), “Socio-economic, ethnic and gender differences in HE participation”. *BIS Research Paper No.186*. BIS: UK, London. Available from: <https://www.gov.uk/government/collections/bis-research-paper>

² Bowes L et al. (2015), “Understanding progression into higher education for disadvantaged and under-represented groups”. *BIS Research Paper No.229*. BIS: UK, London. Available from: <https://www.gov.uk/government/collections/bis-research-paper>

³ House of Lords Select Committee on Social Mobility (2015), ‘Overlooked and left behind: improving the transition from school to work for the majority of young people’, The Stationery Office: UK, London. Available from: <http://www.publications.parliament.uk/pa/ld201516/ldselect/ldsocmob/120/120.pdf>

Currently, the UK longitudinal study community are keen to embrace the potential offered by survey and administrative data linkage but are being hindered by disparate and opaque data access processes and procedures. Our current experiences are described very accurately in paragraph 95 of the consultation. For example, one of CLOSER's projects seeks to link survey data to records held by HMRC (only in cases survey participants had consented). Progress on this has been significantly delayed because of practical difficulties within HMRC in extracting the relevant data from their systems at a feasible cost.

We welcome and support the Cabinet Office initiative and its recognition of the difficulties that researchers face when trying to access administrative data for research of public benefit. Our comments focus on the areas within the consultation which relate to improving access to de-identified administrative records for research purposes (Section C.i.). In summary, we have seven areas of comment on the proposed legislation:

1. It is unclear whether access to administrative records in order to link to survey data would be covered within the legislation; we would strongly urge that such linkage requests are explicitly described as being within its scope.
2. In order to robustly utilise shared records it is important that these are documented to a standard that allows researchers to understand the information and the mechanisms by which the information are collated.
3. We would like to see stronger mechanisms in place to incentivise data sharing by government departments.
4. We believe that greater flexibility should be introduced as regards the mechanisms by which data linkage will be enacted. We would like to see experts from the research community involved in drawing up associated Codes of Practice.
5. We note the new considerable role for the UK Statistics Authority and urge that this is sufficiently resourced in order not to create new bottle-necks in the research process.
6. We support the decision not to define 'public interest' within the legislation.
7. The proposals should clarify whether and how researchers might access health and social care data and use these data in conjunction with survey data, potentially alongside other administrative records.

The following paragraphs expand on each of these points in turn, while also responding to three of the specific consultation questions (15-17).

1. The status of survey and administrative record linkage

- 1.1. As outlined earlier, linking survey and administrative data permits research of considerable public benefit which offers more than either form of data can contribute alone. Currently, attempts to access the necessary administrative data encounter precisely the same problems that are articulated in the consultation. However, survey data linkage is not mentioned in the consultation. We feel this is a missed opportunity and would strongly advocate its explicit inclusion in the legislation. We would be happy to provide further input into discussions on this area.
- 1.2. The proposed legislation would be in line with the type of data protection processes already adopted by individual surveys in order to protect the identity and privacy of people during and after the research process.

2. Documenting shared data

- 2.1. The proposed legislation details the sharing of information in the form of a dataset of records and provides a mechanism by which these data may be shared. We note that in order to effectively utilise these records in a scientifically defensible manner requires researchers to receive sufficient documentation (or 'metadata') to understand the data; the way in which the data were collected, processed and the potential errors and biases that may be present and the way in which all of these factors vary across time (i.e. through the implementation of version control). We suggest that sufficient emphasis is placed on this to encourage data providers to collate robust documentation and to feel empowered to share the documentation along with the underlying records. The way administrative data has been linked also needs to be fully documented and involve best practice.

3. Incentivising data sharing

- 3.1. We welcome the proposal that government departments should share and link more data through appropriate mechanisms. However, we note that the legislation will be permissive in nature, rather than requiring departments to share. We would like to see consideration of stronger incentives to encourage departments to participate. In our experience, there are a number of barriers to successful data linkage, only one of which (legislative uncertainty) will be currently resolved by these proposals. Other barriers include resource (including insufficient staff with the necessary data management skills and high staff turnover) and out of date data management structures or procedures.

- 3.2. The Cabinet Office is consulting on the proposal that departments charge a fee to the applicant, on a cost-recovery basis, in return for supplying data for linkage (paragraph 100, consultation question 15). We are not opposed to proportionate charges, so long as these are modest and their calculation is transparent. However, we are not convinced that all departments are set up to operate on a cost-recovery basis or, crucially, that this approach will incentivise the cultural change required within departments (our views about current barriers to data sharing are set out in the previous paragraph).
- 3.3. We are concerned that high fees may form a barrier to some public interest research and suggest that the UK Statistics Authority consult with the public and charity organisations who fund research about appropriate fees.
- 3.4. We note that, if fees are charged, a mechanism will need to be developed whereby firm quotations can be provided to researchers for grant applications to Research Councils and other bodies, with a commitment to these being honoured if the grant application is approved.

4. The data linkage mechanism

- 4.1. Paragraphs 97 and 98 outline the recommended model for data linkage, the Trusted Third Party model. Paragraph 107 recognises that best practices for de-identifying data are likely to change over time, and proposes that the procedures can be modified by secondary legislation. We believe that scope for revision of the exact mechanisms by which data linkage is enacted is vital, in order to avoid building in obsolescence and a disincentive to improve methodologies. We would argue that the legislation should enshrine the principle of 'effectively anonymised' rather than specifying the exact mechanism by which this should occur. We would like to see research experts and other stakeholders involved in developing a relevant Code of Practice.
- 4.2. Paragraph 97 cites the Administrative Data Taskforce (ADT) report and its recommendation of the Trusted Third Party model. We note that this report did not only provide one data linkage solution as it recognised that other approaches solutions might be required in particular circumstances. We also note that this report advocates data destruction after initial use. We would strongly argue against this approach, particularly if linkage between administrative and survey data becomes within the scope of the legislation. As recognised by the Information Commissioners Office's Code of Practice on anonymisation, certain research (including longitudinal research) requires retention of administrative records for an indefinite period.⁴

⁴ As outlined in Section 9 of the code of practice: <https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf>

5. The role of the UK Statistics Authority

- 5.1. Paragraphs 104, 105 and 108 describe key activities to be carried out by the UK Statistics Authority. These include the accreditation of all active roles within the data sharing process (indexer, access facilitator and researcher), accreditation of the research (including decision-making about whether the proposed research will be in the public interest). We welcome the role of an oversight body but would urge that it is adequately resourced for this function, in order to avoid it becoming a bottleneck.
- 5.2. We note the existence of existing relevant documentation and standards which we would draw to the consultation's attention. In addition to the Information Commissioner's Office code of practice on anonymization (which we touch on in para 4.2), these include Data Safe Haven concepts⁵, ESRC guidelines regarding the access of Approved Accredited Researchers to secure data⁶, and the ISO27001 Information Security international standard.
- 5.3. It would be useful to have clarity as to whether Section 33 of the Data Protection Act (which includes some exemptions to the Act for research purposes) is compatible with any new data sharing legislation.⁷
- 5.4. If linkage to survey data is to be covered by the legislation, we would like to see it being possible for individual studies to operate as the accredited access facilitator (paragraph 202). These studies would be responsible for arranging the secure combination of participants' study and administrative records, and onward sharing these with accredited researchers to conduct accredited research.
- 5.5. The Cabinet Office is consulting on whether the UK Statistics Authority should publish details of rejected applications and the reasons for their rejection (consultation question 16). We support the principle of publishing details of why specific applications were rejected, for transparency reasons, but thought will need to be given as to how much detail about the rejected application should be made public as an application may contain valuable Intellectual Property in addition to confidential information about applicants' information security arrangements.

⁵ Burton et al (2015), 'Data Safe Havens in health research and healthcare', *Bioinformatics*, 31:20.

⁶ <https://www.ukdataservice.ac.uk/media/311608/CD140-SecureAccessAgreement.pdf>

⁷ <http://www.legislation.gov.uk/ukpga/1998/29/section/33>

5.6. As outlined earlier, we have concerns about the extent to which departments will be incentivised to provide data once an application has been approved. Given this, we would urge the Cabinet Office and the UK Statistics Authority to give thought to other information that might be made publically available, and which might act as an incentive for departmental participation. Examples might include the number of applications made to specific departments for data, the number of datasets provided, time taken to do so, and so on.

6. Defining public interest

6.1. Paragraph 108 states that the legislation does not attempt to define ‘the public interest’. Instead, the decision as to whether a research application meets public benefit criteria will be a task for the UK Statistics Authority. The consultation asks for views on what principles or criteria should be used to identify research that has the potential for public benefit (consultation question 17).

6.2. We support the proposal that public benefit is not defined in the legislation. Our main recommendation is that a wide definition of potential public benefit is used, and that public benefit is not linked explicitly to the improvement of public services. Research using administrative records can be of wider public benefit – for example, work on understanding the causes and consequences of social change, or exploring the impact of policy and practice on individuals and their families.

6.3. We would recommend that organisations seeking access to administrative data are required to publish outcomes from this research and place them in the public domain. We note that this commitment to publication forms part of the MRC definition of Bona Fide research.⁸

6.4. Non-public bodies engaging in research (for instance, commercial research companies and third sector organisations) are not clearly covered by the proposals. We would welcome a clear statement about whether such organisations would have access to data. We note recent Wellcome Trust research, prepared to inform the Caldicott Review, which shows considerable public support for access to health records by academic researchers, charities and organisations working with the public sector. It also identifies a significant minority which objects to commercial organisations having access to health records under any conditions.⁹ We note that concern about commercial access to health care records was a major concern in the public reaction to Care.data. Given this, we see it as vital that specific mention is made of the access arrangements and safeguards deployed if different non-public bodies were able to access administrative records.

⁸ See: <http://www.nshd.mrc.ac.uk/data/data-sharing/meta-data-repository/bona-fide-research/>

⁹ Ipsos Mori (2016), ‘The One-Way Mirror: Public attitudes to commercial access to health data’. Available at: www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Personal-information/Public-engagement/index.htm

7. Integration of health and social care data

- 7.1. It is not clear whether and how it will be possible to link administrative health and social care records to other records. We appreciate that the scope of this consultation in this area is very limited (as outlined in paragraph 28), as these data are subject to separate legislation and are under review by Dame Fiona Caldicott. We see it as vital that any standards for research that emerge from the Cabinet Office proposals are integrated with those that are developed for patient records. Data linkage across health and other domains underpins hugely valuable research of considerable public benefit and should not be prevented by departmental differences in data access.

22 April 2016