ESRC Data Infrastructure Strategy

Stakeholder Engagement Feedback

Please answer each question in no more than 200 words.

Completed forms should be sent to datainfrastructure@esrc.ukri.org by 10 September 2021.

1. What do you think about the vision for the Data Infrastructure Strategy? For the vision, see section 4.3.

We agree with the vision and especially commend the focus on innovation, integration and flexibility, all of which are required to achieve the ultimate goal of addressing major societal challenges and advancing the UK's reputation as a world leader in social science. However, the necessary supports for innovation, integration and flexibility – including timeliness and affordability of data access - are currently lacking when it comes to linkage between the ESRC surveys and administrative data sources, and also between administrative sources across sectors and nations of the UK. Unfortunately, it was not clear to us how improvements to any of these will be achieved based on the objectives and theory of change model which are lacking in detail and specificity.

2. What do you think about the activities identified to support delivery of the strategy and do you think any others should be considered? For the activities, see section 4.5.

We agree with the intention, and especially support activities relating to data linkage; discoverability and access; training and capacity building; responsiveness; landscape oversight and co-ordination. The achievements of ESRC are world leading with respect to longitudinal surveys and, more recently, harmonisation across these surveys via initiatives such as CLOSER. A major barrier to realising this full potential has been around linkage to other (administrative) data sources. We believe this is an area that must be advanced in order to meet the stated vision and is one which the ESRC could have a leading influence. However, the activities need to be made more specific to have impact, e.g:

1. Increasing data science capacity: New forms of distance and online learning courses for social scientists to gain skills and knowledge around governance, ethics, analysis. Undergraduate and Masters courses should include data linkage, and administrative data and government data.

2. Supporting the generation of synthetic data to train the next generation of researchers, without the risks (and delays) associated with accessing individual-level data from government departments.

3. Funding calls focussed on early and mid-career expertise in areas where a critical mass is lacking; small seed funding as a springboard to larger applications.

4. Change the conversation from the risks of privacy breaches to the risks of not doing research (e.g. burden of deaths/disease that could be avoided, costs saved from ineffective programmes). COVID has played an important role in the public's perception of the value of timely data use, and could form the basis for a national campaign.

3. Do you have any comments on the draft theory of change (Figure 1), focusing on evidence of need and targeted impacts?

We agree with the stated needs and impacts, but find the activities currently lacking in detail. We believe that the impacts outlined in the ToC model cannot be addressed without triangulation or linkage between longitudinal survey studies and administrative data sources. For example, cohorts do not always lend themselves to evaluations of the upstream policies which are most likely to address some of society's biggest challenges (e.g. reduction of inequalities). And while panel studies provide trend data, they can be limited in statistical power to look at changes of specific policies on specific groups (for example early years' policies on maternal and infant health). Administrative data are well placed and well powered to evaluate the impacts of policy changes over time or between countries, for rare outcomes and within population subgroups, for providing good representation, however they are less well placed for understanding / accounting for contextual effects, the mechanisms through which any policy impacts may arise, and any threats from confounding.

We also note that the uncertainty and long lead times currently involved in data access, and data linkage between surveys and administrative data hamper research impact and in particular capacity building at the PhD and post-doctoral level (making projects high risk and often reliant upon less ambitious and less impactful contingency plans). The activities suggested in earlier sections, including better resourcing, greater engagement with gatekeepers and more flexible funding, could help to address this.

4. What do you think about the draft goals and objectives in Figure 2?

A significant challenge (Goal#1) is maximising the full potential of UK observational data via linkage; this is crucial for achieving Goal#3 (innovative, holistic, comprehensive, ambitious approaches). Surveys and administrative data have complementary strengths and progressing our ability to link them would seem the most productive way to create 'a sustainable mainstream research resource'. Goal#2

(to 'enhance connections with other research councils to explore linkages with other disciplines') must be more ambitious – e.g. to include government bodies which hold data across different sectors (e.g. social care, welfare, housing, education, employment). This would all be best taken forward as a UKRI-wide effort. We found the distinction between 'people' and 'infrastructure' in Goal#2 unhelpful – a common reason behind delays in data linkage and access is the demand for expertise in the datasets, extraction, linkage etc. outstripping supply.

We suggest that the ESRC could help to:

-Advocate for reduced data costs through core government funding (avoiding cost-recovery models which are beyond the means of most data analysts, as in Sweden has done (where costs are subsidised by ~75%).

-Improve infrastructure for data access and analysis through remote systems and drive forward improvements in processing time and data storage.

-Consider how to develop a more diverse and specialised capacity for providing access to linked data, e.g. the use of independent providers may encourage innovation, flexibility and staffing, reducing delays in data access and outputs.

Steps to achieve Objective#4 could include learning from examples of best practice, e.g. SWS, ALSPAC have strong track records of engaging with participants, as well as other nations (e.g. Australia) which have better public perception of data sharing. Delving into other issues such as mistrust in private organisations and legislation (e.g. Scotland and England have differing interpretations of data sharing under common law).

5. In your view, and after reviewing this document, what impact will the strategy make?

Minimal – while we wholy support the vision it is hard to see how these important but ambitious goals will be met via the proposed objectives and plans which are lacking in detail and commitment. We hope that the activities and objectives suggested in earlier sections might help in relation to linkage between adminstrative sources.

We would like to see an investigation of the current barriers and facilators experienced in accessing and analying linked data; the reasons underlying the variation in experiences (e.g. why the MCS could be linked to health records in Wales and Scotland but not England); and what is required to overcome these wholesale (this has been achieved for COVID but barriers remain widespread elsewhere). Consideration of the costs of these missed opportunities might help to make a case for improvement in the current system – including researcher time spent waiting, chasing and justifying linkages, multiple linkages of the same datasets being carried out by resource-stretched governmental organisations, data being collected in surveys which is already available routinely. Learning from the experiences of other countries who have been more successful in this regard would be valuable. Data scientists are likely to experience similar barriers to progression in academic careers as statisticians (as has been highlighted by the Royal Statistical Society). We would welcome greater investment in career structures for data scientists in academia, especially at earlier career stages.

6. Is there anything else that you would like to comment on?

The Maternal and Child Health Network (MatCHNet) is a UK Prevention Research Partnership (UKPRP) funded network that aims to harness the UK's rich administrative data sources and the differences in the timing, eligibility or focus of policies across the 4 UK counties, to understand policy impacts on child and maternal health inequalities. Our members span disciplines (e.g. public health, epidemiology, economics, statistics and data science), sectors (e.g. academic researchers, data holders and processors, policy-makers, health practitioners and advocates for child and maternal health) and the four UK nations.

7. Would you be happy to be contacted to discuss your comments and would you like to be involved in any additional engagement activities?

Yes