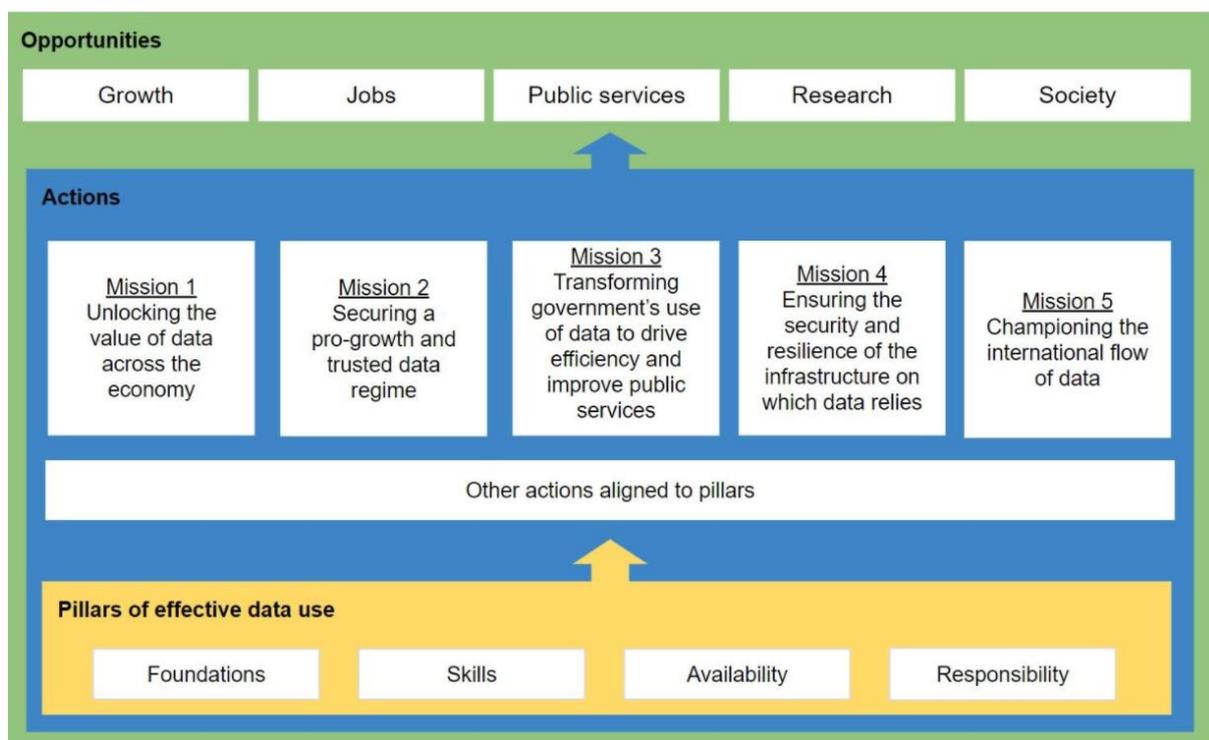


## ACADEMY OF SOCIAL SCIENCES RESPONSE TO UK NATIONAL DATA STRATEGY CONSULTATION

### Introduction

In September 2020, the UK government published a consultation on the UK’s [National Data Strategy](#). Led by AcSS Council member Dame Jil Matheson, a former National Statistician, and working with some of our Learned Society members and others, AcSS submitted a response in early December 2020. Because of the particular structure of the consultation questionnaire, which would have required reading the full consultation materials, we agreed to publish an ‘outward-facing’ version of the Academy of Social Science response.

The National Data Strategy is underpinned by a model based on five ‘missions’ (the aims of government action) and four ‘pillars’ (the things that need to be in place in order for those missions to succeed). This can be summarised in the diagram below, taken from the National Data Strategy.



## Some general observations

While we agreed that the five missions and the four pillars provided a strong analytic structure, AcSS observed that many of these actions and descriptions were treated in relatively narrow technical terms, without appreciation of the links between them or the broader, medium-term institutional actions that would be needed to achieve the government's ambitious aims.

The four pillars are: **foundations** – ensuring data are fit for purpose; **skills** – for a data-driven economy and data-rich lives; **availability** – ensuring data are appropriately accessible; and **responsibility** – driving safe and trusted use of data.

For instance, there was little discussion of the links between data foundations – whether the data were truly adequate for the task, and therefore use of them was trustworthy -- and availability and responsibility – how to set up an appropriate regime that balanced public benefit in the use of data with a governance regime that was both trustworthy and trusted. We believe social science practices and expertise has much to contribute here. And we noted that there were no separate questions on data skills, which were again largely technically-defined.

Instead, the National Data Strategy set out a number of mainly technical and resource-based considerations relevant to the individual pillars, covering issues such as curation, technical data transfer protocols and so on. These are undoubtedly important. But the institutional practices that support the pillars are also important.

We summarise our responses to the questions under each of the four pillars in the consultation document.

## Data foundations

The National Data Strategy set out a question about government responsibility for **data foundations**:

“The true value of data can only be fully realised when it is fit for purpose, recorded in standardised formats on modern, future-proof systems and held in a condition that means it is findable, accessible, interoperable and reusable. By improving the quality of the data, we can use it more effectively, and drive better insights and outcomes from its use. To what extent do you agree that the government has a role in supporting data foundations in the wider economy.

In our response, we pointed out that technical standards, curation of data and so on are undoubtedly important, and that others were better qualified to comment on the technical aspects. But we pointed out that data adequacy required more than this, including statistical understanding of representativeness and generalisability, and substantive understanding of the limits of inference in particular circumstances to do with the properties of the data, and the issue its use was intended to address. These issues – the limit of inference and appropriate use of data, and substantive understanding of the data in relation to the problem at hand – would be at least as important as technical features of data in making specific decisions about data availability, including promoting wider use, and government support for these. We also noted that the accountability and government framework (both legally and

institutionally) would be important, and should be considered as resources were deployed by government to create data foundations, and decide whether and how to make them available. The Academy of Social Sciences believes that some sort of triage model would be needed, with different data in different circumstances and sectors, in deciding how to resource data foundations and link data availability to risk (to society or individuals), data governance, and public benefit.

The consultation asked two specific questions relevant to data foundations.

First, it asked for “examples of how data was or should have been used to deliver public benefits during the **coronavirus (COVID-19) crisis**, beyond its use directly in health and social care’. It was explicit that health and social care data would be covered in a future data strategy for health and social care.

In our response, we felt we could not ignore the issue of access to health and social care data. We argued that the [COVID-19 Hub](#) of the Campaign for Social Science (part of the Academy of Social Sciences) shows just how wide-ranging the topics are that benefitted from data use during the first six months of the coronavirus crisis. Indeed, many of the important responses, not just on our hub but also in government policy-making and implementation, were precisely those requiring social science evidence and data, including: educational inequalities; labour market conditions including sectoral understanding; work relevant to the ‘gig’ economy; the distribution of differential household financial resilience; transport use; social behaviour in a range of sectors and settings; the need for greater geographic and demographic granularity in data, which is best generated from well-structured, standardised and appropriately accessible (privacy-protecting) data.

It is clear to us that the difficulties in linking health and social care data with a wide range of social and demographic data has been a real stumbling block to understanding COVID-related matters as important as the higher incidence, mortality and morbidity of BAME communities in the UK. It is clear that much of that higher incidence was a result of different occupational, household compositional and other socio-economic and behavioural features of different communities within the UK.

The Academy of Social Sciences believes that the governance and security aspects of the UKRI-ESRC supported Administrative Data Network provides a secure, trustworthy and ethically-sound model for achieving wider data access (as we discuss below) and would like to see an end to, or a integration with, the separate regime for health data. We understand the background to the current system in use for allowing accessibility to health data and the reasons for it, but we believe it hampers the UK in addressing a range of issues where health outcomes are linked to social and behavioural features.

A second question asked about **regional inequalities**. We agreed that this was an important issue.

In our response we wrote that one of the most pressing needs is for better data about a range of geographic areas: from small area statistics and travel to work areas, to various local and regional local and health authorities, to the devolved nations. By ‘better data’ we mean more granular data, data than can linked with other sources, and that can be used to compare areas for similarities and differences. The creation and curation of such datasets,

and providing robust governance, privacy safeguards and so on are all a challenge, and perhaps not best led only by central government. Yet the data analytic capacity of many local and regional authorities has become weaker over the past decade, rather than stronger.

We suggested funding regional and devolved hubs for such work, perhaps with regional universities involved, may be a useful way forward. The University of Essex for instance plays some of this role with regard to the local authorities near it; the University of Manchester has been involved in work to link local authorities and local health authorities around the greater Manchester area with good and granular data; and the devolved hubs of the Administrative Data Network in Scotland, Northern Ireland, and Wales, have also shown what such a model can achieve.

We pointed out too that addressing this issue would also require attention to the regional distribution of data skills, as it would promote more equal regional distribution of those who possess higher and technical data skills and capacities.

### **Data skills**

There were no questions about data skills in the consultation. Instead, the National Data Strategy sets out the need for a wide range of data skills and promises to consider them more fully in the forthcoming Digital Strategy, and the next stages of the R&D Roadmap, including providing more maps of types of skills and ‘pathways into data-related careers’. We discussed this issue in our response anyway.

The Academy of Social Sciences welcomes the acknowledgement that a wide range of skills is needed. But these need to be seen as providing a dense ecosystem to promote wider data skills for most, including skills in data collection, data interpretation and analysis, and a range of more advanced skills, some more purely technical, and others more advanced skills in the data particular to various domains, subjects or from certain sources. Only by considering the potential for data skills to transform most sectors and ways of working, and providing the educational underpinning of this, will the government’s aspirations be achieved.

If this is to yield significant long-term results, it requires thinking about its implications throughout the educational life-course. Building in positive attitudes towards numbers and data would need to start early, yet we know that much primary school teaching is done by teachers who do not have special skills in this area, and may themselves feel unconfident. This is not just about teaching primary maths, but thinking about how to build interest in using data across a range of subjects and levels, to lay the foundations for a later learning. The Department for Education should review this, and consider how to address it: in the national curriculum, in the development of teaching materials, and in teacher training, including CPD, and recruitment. Central government has a key strategic role to play in bringing together the various stakeholders to effect real change.

There are more concrete steps that can be taken in the relatively short-term to create more varied pathways to this in secondary schooling – well before the higher technical or higher qualifications that are mentioned in the NDS. Only a wholehearted engagement with

pathways in and from secondary schooling will provide the expansion of the feeder pools that will be needed, as well as providing subject-specific data skills across the UK economy.

We argued that no single pathway or qualification will meet the needs of the UK, both for the highly technical skills that are needed, or more widespread ability in extracting meaning by using data.

In England and Wales, this would mean:

- Bolstering the large-scale data module within the existing Maths A-level, and providing much better and more engaging teaching materials, and supporting the efforts of Ofqual and curriculum-awarding bodies to develop appropriate materials, including support for teachers both in using them and in assessing exams. It was clear from 2019's exams that many 'maths' teachers are still uncomfortable teaching data-led materials, including its statistical content and the need for inference and interpretation.
- Considering bringing back Maths AS level, which would include the large-scale data module, for those students who would benefit from it, but don't want to go on to the second year of the current A-level. This was a qualification that had seen marked growth in uptake in the decade before it was discontinued as a free-standing qualification. Working with UCAS to award points for those who took this particular 'half an A level' would be an important stimulant to data-skills, irrespective of views on the desirability of broadening A-levels more generally.
- Developing a new data science curriculum and qualification at A-level. A promising candidate, worked out in some detail, is already at hand from the [International Data Science in Schools Project](#), supported by the Royal Statistical Society and others.
- Considering significant investment in, and policies to, recruit teachers who have the skills and flair to teach data skills, in a broad sense. This is not just a matter for the DfE but should be a cross-departmental initiative.
- Bolster the data content of various other A-levels. This would require a step change in curricular content and assessment of it, in a subject-specific way. Research conducted some time ago by the Nuffield Foundation showed how variable the 'number and data' content of many A-levels were, both between subjects and in the curriculum set by different awarding authorities. But positive change is possible. The Royal Geographic Society, for instance, has shown that the number and data content of an A-level can be significantly enhanced, and appropriate teacher support provided. The psychology A-level has taken some steps in this direction. We believe a range of subjects, as diverse as biology and chemistry within the STEM sciences, as well as various social science disciplines could benefit from this approach.
- In addition to providing a larger and more diverse pool of young people who might then be interested in going to more advanced data science work, this would also tackle issues such as diversity (many social sciences, and, within STEM, biology, are, for instance, taken by a larger number of girls than boys).

- Finally, more can be done across a range of higher qualifications, including technical apprenticeships (as the NDS notes) and even within university courses. For instance, the Nuffield/ ESRC Q-Step programme produced a real step change in university teaching capacity and student numbers using data in the social sciences in the universities that took part. The Academy of Social Sciences has suggested to UKRI - ESRC that it should increase the training in data-handling and use at the postgraduate level, but that doing this would benefit greatly if steps were taken to do this at undergraduate level too.
- Our recent report, [Vital Business](#), shows how private sector businesses value data-skills are in social science graduates, who perform a number of key functions in these businesses, as well as work with STEM scientists and others in innovation chains, scenario and risk planning, and so on.
- So the UK needs to address a short-fall not just in technical data scientists and AI specialists, though it undoubtedly needs those, but also to create a much wider and deeper pool of those who can conceive and use data science more broadly. Meeting this need would take a real concentration of efforts and resource to achieve the step-change that is needed if higher-level technical and occupational qualifications and centres are to take off.

An emphasis on data skills were also to the fore in our response to a question about what central government could do to tackle the barriers that small and medium-sized enterprises (SMEs) face in using data effectively. We noted that barriers that exist for SMEs include not only data skills and capacities within the SME, but also general capacity to consider such issues in the face of challenges of running or working within an SME. Many simply do not have the scale to generate the bandwidth or bring in the skills. We argued that central government could not only enable but resource resourcing intermediary bodies who can provide data-based evidence. There are examples in some of the work done by local authorities or local chambers of commerce, some developed under the UKRI-ESRC ADR, and others under UKRI's 'Strength in Places' partnerships. Innovate UK could also play a role, funding capacity for SMEs in particular sectors or regions, at least with seed-money. We also thought central government could support local authority capacity, with public-private partnerships, to service some of this work in important local sectors would be possible under the place-based agenda and budget run from BEIS, not UKRI, though local universities may be useful partners in some cases.

### **Data availability and responsibility**

The AcSS responses to other questions emphasised the essential links between **data availability** – ensuring data are appropriately accessible; and **responsibility** – driving safe and trusted use of data.

For instance, our response to a question about the potential impact of the proposals on **individuals with a protected characteristic under the Equality Act 2010**, highlighted this issue.

We noted that AcSS believes that the general features of a National Data Strategy that are important for individuals with protected characteristics are also important to all citizens. That will require in the main:

- Attention to the content, collection, structure, curation and generalisability of data, with clear scientific review and justification and a public benefit test in which there are independent voices from outside government;
- Clear and transparent governance in making the public benefit justification, again with independent voices from outside government;
- Robust protections for privacy, including not only technical protections but consideration of who end users are and how their behaviour can be incentivised and sanctioned (if need be);

We argued that only by ensuring all these are in place will there be a trusted data regime – fundamentally because the regime is trustworthy. We note that this not only requires open publication of data registers, but shows how the pillars of the National Data Strategy are linked. Trustworthiness is not a product only of ethical review but about making sure the pillars support each other. Having truly independent representation – among those who scrutinise proposals and those who use data – from outside government or business end-users is important to strengthen trust, and trustworthiness.

We added the data linkage and data sharing that is likely to be required to make a positive difference to individuals with protected characteristics is often likely to be particularly sensitive, and we note that, as with some previous government data initiatives, trust can be quickly lost. Building a robust system across the piece is the best safeguard.

**The Academy of Social Sciences believes that some sort of triage model is inevitable, with principles of open transparency and accountability, data security and access requirements operating across all areas, but with different degrees of scrutiny and potential sanction, depending on the risks to society or individuals. In other words, within the principles set down by a principle-led regulatory regime, there should be robust risk and benefit triage.** This can still be accomplished in an agile way, as the UKRI-ESRC supported Administrative Data Network shows.

We also noted that the [Wellcome Trust's research on data sharing in health](#) (and [here](#) on commercial access to health data) has strong lessons in this area. We are strongly of the view that independent voices, perhaps nominated by the Data Commissioner, or Turing or a number of other relevant foundations or organisations, or (as in the case of the Smart Data question, by relevant sectoral consumer-groups) are essential. This provides a stronger public benefit test so that commercial interests have to be transparent in justifying any public benefit test, and in helping approve relevant governance. This is, we believe, essential to ensure trustworthiness. Government control or approvals alone are unlikely to achieve this.

We returned to the importance of scrutiny by those outside commercial interests or in government in assessing public benefits tests, and in having access to data as independent users. For instance, we responded to a question about “**smart data**” -- so that in areas

such as banking, finance, telecoms and energy, data about citizens could be shared with third parties, by highlighting the issue of trustworthiness.

We acknowledged this data sharing regime has had benefits both to companies and, more importantly, to citizens and consumers, allowing more consumer-friendly regulatory regimes and making it easier, for example, for consumers to switch providers in these sectors.

We argued that a key element in these data-sharing schemes has been engagement with relevant consumer organisations and regulators, as well as the firms themselves. Allowing access to some data, both aggregate data and appropriately protected individual data, by regulators and relevant consumer organisations is a key to ensuring consumers reap benefits, and that they trust the data-sharing regimes because it brings genuine benefits to them. Any similar new schemes should take care to involve regulators and consumer organisations in the governance and scrutiny mechanisms, and in data use.

We made similar points – about the linkage between data governance and data usage, and the importance of requiring appropriate involvement of those outside government in both, if trustworthy data regimes were to be built – in our response to a question about the **UK's data protection framework**.

We noted that Britain's General Data Protection Regulation (GDPR) has now settled in cultural landscape, and provides a useful and important regulatory framework. We agree with some of the points in the NDS consultation that some of the barriers to wider data use were cultural and institutional, and a tendency to over-cautiousness in the context of legal and other risk, but we thought that was now reducing, as the public benefits tests in GDPR were better understood. The Information Commissioner has, in our experience including over the Administrative Data Network and its governance framework, been constructive and helpful. Further work to deepen the cultural change, not least within government, would be helpful.

So we stressed that an important issue was, at least as far as government data was concerned, something that government itself could affect. We noted this would require acknowledgement that achieving trustworthiness requires open and transparent involvement of independent voices, both in governance of, and gaining access to, appropriately secured data – for instance in government encouraging independent academic having access to government administrative data under appropriately-controlled privacy and security settings (as with the ADRN), consumer organisations being involved in Smart Data initiatives, and so on. Data sharing which only allowed those in government departments only, or between government and the private sector only, or between private sector organisations only, carries not only risks to individuals, but risks of more substantive 'capture' that promotes favoured interpretations and uses that may not have wider social backing, or that raise privacy or distributional concerns which are not visible if other experts who can provide public benefit are not given access, or a role in data governance.

We argued that this illustrates the importance of taking the pillars, and the links between them seriously. Open scrutiny of governance, involvement of independent voices, and ensuring appropriate wider access to ensure constructive challenge, both for better data collection, and for different interpretations and data-based inference is important for

generating and sustaining public trust in data use and sharing in the longer-term, however inconvenient it may seem to government in the short-term.

We also noted that if government focusses only on enabling access to data by central government, it is likely to undermine data collection (outside experts are often important in constructive challenge to existing data collection, and to the limits of legitimate inference from particular datasets), and to undermine, potentially fatally, public trust. We added that while this required resourcing, it also required real Ministerial commitment in accepting that building a regime the public trusts requires external challenge and access, whatever its short-term inconveniences at times.

We highlighted other ways that government could promote more transparent use of data and governance scrutiny. For instance, in our reply to questions about **the functions of the Centre for Data Ethics and Innovation (CDEI)** and its role in artificial intelligence monitoring, piloting and testing, we noted the wider landscape available to promote informed discussion, consideration of social features and public benefits arguments, and transparency.

We noted that previous work by the CDEI shows the potential power of bringing together an understanding of technical issues and analytic techniques with social science understanding (of behavioural, and legal and other institutional issues) with ethics and governance involving transparent oversight and accountability mechanisms, as it did with its [2020 report on the use of algorithms for individual decision-making](#). We did not favour CDEI's remit in AI, or other technical aspects, to impede it from considering these wider aspects.

We also mentioned other institutions who had a role in considering these issues, both in policy and in practice in individual cases. These include: the Information Commissioner, the UK Statistics Authority and its Office for Statistics Regulation, and various other regulatory bodies. We believe that any National Data Strategy will be strengthened by recognising the wider institutional structures and system, and the value of having independent expert organisations involved. We added that one of the strengths of the UK Statistics Authority role in supporting good practice and policy in the National Data Strategy arises because it produces an annual report openly to a Parliamentary Select Committee, rather than only to the government executive. We suggested ensuring such a statutory duty to report directly to Parliament in other cases too. We also suggested that in light of the ease of data transfer and technical issues in data security, external experts (from outside government and data users) should be more used more often in assessing data security, and, in a response to question about data security protocols during the life of any contracts, should be accompanied by greater attention to *post-contractual* obligations (curation, data transfer, and data security).

The last questions in the consultation were about **international transfer mechanisms** – specifically, how these could be improved while ensuring that the personal data of UK citizens are appropriately safeguarded.

Here, in common with other respondents to the consultation, we raised some concerns. The Academy of Social Sciences understands the potential benefits of data sharing that might involve international transfers. But any such regime needs to build in the standards at least

equivalent to GDPR, which is a higher threshold in many ways to those in use, both in practice and as a matter of public regulation, in the US or a range of other countries.

We noted that, as well-illustrated by the Cambridge Analytica case, the UK government has much less reach to regulate or sanction multinational actors than it does UK public and private-sector bodies. We argued that in any transfers, the standards set by the GDPR, and by the Digital Economy Act 2017 should not be diluted, and should serve as a 'floor' set of standards. As a matter of practice, we also suggested that independent expert bodies, such as the Information Commissioner, Turing, the UK Statistics Authority, the Royal Statistical Society, and regulatory and consumer bodies should be involved in any concrete proposals for and decisions about, international data sharing.

The consultation said that the UK would seek **EU 'data adequacy'** to maintain the free flow of personal data from the EEA and we will pursue UK 'data adequacy' with global partners to promote the free flow of data to and from the UK and ensure it will be properly protected. In our response, we noted that (at the time of writing), EU 'data adequacy' was being sought, though it had not yet been achieved.

Regarding transfer to non-EEA countries, we argued that, for the range of government and commercial data with which we are familiar, including those infrastructures developed in the UK for data sharing, and based on our understanding of the wider social and economic policy landscape, the standards set by EU 'data adequacy' should be set a 'floor' for any UK rules, and hence any data-sharing arrangements with other countries. Any departures should only be where higher safeguards were imposed.

Substantively, we noted that for many issues, EU country comparators are important because of shared attitudes and behaviours relevant to data collection, sharing and use, and that the regulatory regime in the UK is generally supported; it is already more akin to the protections provided by the EU than the regime in place in the USA. This is made more complex by the federal nature of much relevant US regulation.

We noted the proposal in the National Data Strategy that the UK government could develop its own data adequacy assessment standards, but are concerned that this could, if handled as a matter of closed government protocols, without Parliamentary and wider involvement of non-governmental actors, result in undermining public trust in the uses and benefits of any data-sharing regimes more generally. The UK could productively work with international bodies to set open and principle-led mechanisms with appropriate external scrutiny and sanctions to build regimes based on existing EU principles as a floor.