Royal Statistical Society (RSS) response to DCMS Consultation on National Data Strategy

Preface

The Royal Statistical Society (RSS) is a professional body for statisticians and data scientists with over 10,000 members around the world. Our strategic and charitable goals focus on the use of statistics and data in the public interest, teaching statistical literacy, developing professional skills and the strength of statistics and data in academia and research.

The RSS welcomes the opportunity to respond to the Department for Digital, Culture, Media & Sport (DCMS) open call for evidence for the UK government's National Data Strategy (NDS).

Please note that the tight timeframe for developing a response means that we are submitting a summary version highlighting a limited number of key points. The deadline means we have not been able to consult as widely across our fellowship as we would have wished. The RSS would urge the DCMS to consider the involvement of our expert members in any forthcoming discussions.

The RSS has not responded to the detailed consultation questions but, instead, we begin with some general points, and then consider issues by applying your provisional set of objectives regarding people, economy and government.

We look forward to a full consultation on the NDS later in the year.

General

The RSS welcomes that the DCMS has launched an open call for evidence for the UK's National Data Strategy. The provisional set of NDS objectives, 'people, economy and government', complement the RSS *Data Manifesto* which calls upon the UK government to improve data for 'democracy, prosperity and policy-making'.

The RSS would emphasise that there is already an enormous amount of activity concerning the 'data agenda' across government, as well as in the private sector, universities and civil society. An important starting point is to ask how an NDS emanating from DCMS can add value to the wide range of work already going on in relation to data both within government and outside of it. We would suggest that the strategy can help by:

- showing how existing initiatives and programmes fit together;
- pointing to any significant gaps;
- finding ways to make the whole of this existing activity more than the sum of its parts.

The RSS's *Data Manifesto* set out a number of themes which should inform an NDS, and this informs the views we put forward below. We also commend the UK Statistics Authority (UKSA) for its work in producing a *Code of Practice for Statistics* which has three pillars: trustworthiness, quality and value. These should underpin a National Data Strategy. The National Statistician is a key figure within

government: a permanent secretary with a sole focus upon data. Therefore, the NDS ought to highlight the importance of this role, and see the National Statistician as a key co-ordinating figure.

In recent years there has been considerable interest within government around artificial intelligence (AI). The RSS has been an active participant in these discussions, including through our submissions to consultations on AI and on algorithms. It is critical, however, that the NDS does not focus purely upon AI: there are many data issues which are not specific to AI and, indeed, pre-date the current interest in AI – for example how we can improve the use of evidence in government.

People

1. To ensure that data is used in a way that people can trust.

The RSS agrees that ensuring trustworthiness is an important cross-cutting issue. As data is increasingly ubiquitous and cross-cutting, this is an enormous agenda. The challenges in health will differ from those in education, which will differ again from, for example, the application of technologies such as facial recognition. Such wide-ranging areas are being considered by a range of RSS specialist sections - from Medical to Data Science to Law, respectively. A fundamental way to think about the issue is in terms of governance and accountability. Does the particular application have the requisite governance and accountability – through a mixture of the application of law/regulation, ethics/codes of conduct, and broader professional norms by data professionals?

Considerable work is being undertaken by a range of bodies in this area including the Centre for Data Ethics and Innovation (CDEI), the Information Commission's Office (ICO), the National Statistician's Data Ethics Advisory Committee, the Office for Statistics Regulation, the Ada Lovelace Institute, the Open Data Institute and DotEveryone, while the RSS Data Ethics special interest group is taking an active role in informing this work. A key challenge in this area is for civil society bodies to be able to do independent work to help represent the public interest. The RSS would recommend that the Office for Civil Society (OCS) consider setting up a fund for charities to do work to protect the public interest on data issues.

The ICO is an increasingly crucial regulatory organisation as ever more bodies are using personal data. The RSS has been pleased to see it has been ever proactive in its work. The structure of the ICO means that much pressure is put on the individual Information Commissioner. The RSS would suggest the governance of the organisation is reassessed with a view to creating a full Board that is responsible for the organisation. Such a Board would take some pressure off the individual Information Commissioner, as well as strengthening the ICO's governance.

Given the plethora of bodies in this area, consideration should also be given to rationalisation. The RSS would suggest making the CDEI a part of the ICO. In our view, this would strengthen both bodies: it would provide the ICO with a more extensive horizon scanning and thought leadership function, while giving the CDEI a more permanent home - rather than expecting it to be a small stand-alone body which may find it hard to survive in the longer run.

Professional bodies have a role to play in maintaining trustworthiness. The RSS has a code of conduct which its members should take account of. Our various member groups discuss what makes for good use of data. We are working with the Institute and Faculty of Actuaries to develop ethical guidance for

data scientists in our membership. We also play an important convening role bringing together practitioners, policy-makers, subject matter specialists, etc.

The practice of pre-release access (PRA) to statistics, whereby some people (in practice, usually within the government and devolved administrations) have early sight of statistics, is a breach of ethical and democratic principles. The Office for National Statistics (ONS) commendably ended early access to economic statistics in mid-2017. Furthermore, the Scottish Government in May 2019 *removed* the *maximum of five days* PRA. We suggest this is now extended across all statistics in all parts of the UK.

In keeping with ethical principles surrounding data, we would suggest a national strategy should be introduced by the Office for National Statistics (ONS) focussing on public engagement. This engagement strategy from ONS to users would communicate what data is collected and used to produce statistics and who will have access to it. Furthermore, whilst trust in producing statistics is neutral amongst the public, trust in politicians, governmental organisations and the media to use data properly is a significant issue. It is not good enough to assume that proficient technical methods to protect data are enough, leading to the ONS and the next National Statistician to pro-actively communicate trustworthiness to the public through an engagement strategy. We have given more detailed recommendations relating to strengthening the statistical system in our *recent evidence* to the Public Administration and Constitutional Affairs Select Committee.

Currently, debates are concentrating on whether the right statistics are being collected and if the UK's systems are sufficiently agile and flexible to both develop new measures and to be able to drop what isn't needed/justified any more. So much is published without some idea of uncertainty assessment/statement that it causes problems. According to *Q2.4 from the Code of Practice*, relevant limitations arising from the methods and their application, including uncertainty, should be identified and explained to users. Adhering to 'Q2.4', whilst following the suggestions listed above, should improve the value and significance of the data being collected.

If it is decided that the NDS should focus on measuring public trust, this will be difficult to achieve through surveys, especially as topics such as data sharing are unfamiliar to most people – therefore good survey methods will need to be employed.

2. To ensure that everyone can effectively participate in an increasingly data-driven society.

A critical aspect of this agenda is to ensure that data reflects people's lives. Currently, many official statistics are dominated by national averages. It is often difficult for communities to access data about their locality or region in a way that is useful for them. The RSS encourages the creation of tools to enable easy access to quality data about a given neighbourhood, locality or region. As well as moving beyond national averages, the RSS suggests that official statistics need to give better breakdowns beyond aggregates for multiple groups in society – e.g. across gender, race, age, disability, etc. This way people can have access to the information that they want, in a format that is easy to use.

To ensure that everyone can effectively participate in an increasingly data-driven society, everyone needs to be equipped with the skills to engage in an 'informed data conversation'. The RSS envisages a future with increasing focus and content on statistics and data in teaching practice and assessment across the curricula. This is essential to meet the needs of individual learners, employers and the economy.

The 'data' agenda has many participants, using language in different ways, which means people can talk past each other. It would be useful to start developing common definitions across communities.

<u>Economy</u>

3. To ensure that all businesses and non-profit organisations can effectively operate in an increasingly data-driven economy.

Perhaps the critical issue here is the question of how to improve data skills across the UK. The most significant lever is, of course, the education system. In the RSS's *Data Manifesto*, data is viewed as a critical driver of prosperity. As part of this vision, to effectively operate in the data economy and to upskill the nation, we call upon basic data handling and quantitative skills to be integral parts of the taught curriculum. It is the RSS's view that developing the future pipeline requires a re-think of the current education system to ensure people are 'data-ready' at school and beyond. Building the capability and capacity for people to gain data skills should be an integral part of the NDS. The RSS would recommend that the Smith Review of Post-16 Mathematics is implemented in full by the Department of Education.

Following on from the Smith Review, the RSS encourages the recognition of the need for 'A People Pipeline' as identified by *The Mathematical Sciences People Pipeline Report*. When there is rapid growth in a new subject-area (e.g. data science), it is vital to shorten the delay between the needs of employers and the availability of trained people. As such, it needs to be considered that the education system is a continuous pipeline, established from early years to postgraduate training, which leads to life-long development.

This concept was initially identified by the late Sir Gareth Roberts in his *report* detailing the need for changes in the supply of people with science, technology, engineering and mathematical skills. This was succinctly summarised in the *Kingman Speech*, when Sir John Kingman, Chair of UKRI, stated the following;

> *"we will not have more postgraduates in STEM unless there are more high-quality undergraduates studying STEM. And we will not have more undergraduates studying STEM unless we have more high-quality school students studying STEM at A level. And - particularly since we persist in running in the world's most uniquely specialised post-16 school system – children need previously to have been taught science and maths well enough, and inspiringly enough, to GCSE level to want to go on to take STEM subjects at A level. This is the talent pipeline."*

Different sectors of the economy will need to consider how to improve the data skills relating to their area. We believe these data skills will consist of a mix of traditional statistical knowledge, computer science/coding skills, and domain knowledge of the specific sector. There is a significant question about how these skills will be scaled up through, for example, the NHS, local authorities, the education system, etc. The NDS should encourage each major sector of the economy and public service to consider how they will tackle this.

One way data skills can be taught is through relevant employer training to non-specialists. This will enhance adult employees' capabilities in the workforce which, in turn, will adapt to meet the needs of an increasingly data-driven economy. At the RSS, we recognise the importance of such training. In

doing so, the RSS delivers public training courses to equip non-specialists with foundation skills in topics such as Data Visualisation, Presenting Data, together with Understanding and Analysing Data.

One particular area in which the RSS has developed experience is in working with not-for-profit organisations. Through the RSS's Statisticians for Society *pro bono* programme, we have linked over 50 non-profit organisations to statisticians who volunteer their time free of charge. This programme has made us aware of the low level of data skills in the not-for-profit sector. Again, the OCS should consider a fund to strengthen that sector's data skills.

4. To improve growth and productivity through the effective use of data across the economy.

Good quality information drives economic, social and environmental progress. In the UK, official statistics are critical to this. The RSS believes an essential requisite for data, aside from the need for access by users, is that the data is accurate and fit for purpose. It is therefore imperative for the growth, productivity and advancement of any economy, that the data collected is produced using sound statistical methods. As we move towards a greater use of administrative data or other non-survey type data, this will require the application of statistical methods to ensure the data can be used meaningfully.

In any upcoming Comprehensive Spending Review, it is critical that the nation continues to invest in high quality official statistics, including money to innovate as new data sources become available. The work of the Geospatial Commission will also be valuable as a source of growth and productivity. Data should be thought of as an infrastructure issue for the economy in the same manner as – for example – roads, bridges and electrical grids. Where possible data should be made openly accessible according to open data standards.

One risk to the economy is from potential private sector 'data monopolies'. Network aggregation effects that come from data mean that as organisations hold more and more data, it can be difficult for other organisations to compete with them. RSS supports the recommendations of the *Furman Review* and, in particular, the need to reconsider competition policy to recognise that there may be monopolistic issues even when there is not apparent or immediate consumer harm because services are provided for free of charge.

Government

5. To improve public services and government operations through the effective collection, sharing and use of data.

The RSS would echo that the government should value data as a priority: in summary, data matters. We commend the National Audit Office report on *Challenges in Using Data Across Government*. It sets out a useful framework that the government should take on board in order to make the most of data. The RSS emphasises that whilst there is a need for data to be utilised and shared safely and appropriately, there is also a need to invest in improving and using data for better outcomes.

Building the capability, leadership and culture for evidence-based decision-making requires investment in the skills set across the public sector and government operations. Improving the effective collection, sharing and use of data necessitates that politicians, policy-makers and

professionals in public services (such as regulators, teachers and doctors) should be given training in data handling and statistics. This is something we call upon in the *RSS Data Manifesto*.

The RSS would like to see a much greater focus on the use of evidence to inform policy-making. We accept that policy is not just about evidence – it is the outcome of democratic processes, ideology, and a wide range of other factors. Nevertheless, we would argue that all policy announcements should at least indicate the evidence base in that area, and how far it was taken on board. One gap in the 'evidence infrastructure' is the translation of academic research into evidence that can easily be grasped and used by practitioners and decision-makers. UK Research and Innovation (UKRI) should consider spending more of its budget on such translational activity, including knowledge synthesis so that more publicly-funded research becomes accessible and useable. UKRI should also develop a cross-cutting strategy for data and statistics as this is a cross-cutting issue across science.

6. To achieve alignment in government around data, with data shared and used cooperatively wherever appropriate.

One of the most significant blockages within the government at the moment is around data sharing. Despite the Digital Economy Act, several issues remain. Leadership is required within departments to see the value of data linkage and to invest in this. The ONS will need proportionate and duly justified access to a wide range of administrative data in order to produce the figures which are needed as a nation. This is particularly true for the 2021 census. Additionally, the government needs to get better at sharing administrative data with researchers through initiatives such as the Economic and Social Research Council's Administrative Data Research Partnership. Finally, the bifurcation of health data and other administrative data is unhelpful; the Digital Economy Act covers the latter but not the former. RSS would like to see these regimes brought together over time, to facilitate applications between social and health data.

Data availability is transforming statistical practice by making more inquiries possible, and by removing some of the gatekeepers for accessing data in conventional statistical practice. It is easy to access data without partnering with teams with relevant substantive expertise, but the capability of established organisations to review the ethics of applications using administrative data is limited. Technical, substantive and governance capability will need to evolve rapidly but currently, lack forums for cross-fertilisation. This evolution is most likely to be achieved by convening professionals and their representative bodies and encouraging practitioners to seek suitable accreditation.

Issues surrounding data being shared and used cooperatively was identified in the *Independent Review of UK Economic Statistics, by Professor Sir Charles Bean*. Section 4.38 criticised ONS behaviours and capabilities, stating those within the ONS were 'operating in silos'. To achieve alignment in government surrounding data, the RSS would like to see the implementation of the relevant recommendations in this report. Following on from this, an update should be provided for all those concerned (e.g. RSS and other bodies), confirming the progress on issues as identified within the report and the requisite steps that will need to be taken going forward.

Overall, the Royal Statistical Society believes that a National Data Strategy needs to be ethically founded, while conceived and maintained with quality statistics. Additionally, it will be imperative to provide education and access required by the citizenry, private sector companies and Government to make the National Data Strategy functional. The RSS offers the use of its members who hold a proficient understanding of statistics, to make the NDS an unequivocal success.

Appendix

McConway, K. (2013). '*Remarks on the Future of Academic Statistics in the UK*'. London: Royal Statistical Society. Available at: https://www.rss.org.uk/Images/PDF/influencing-change/RSS-report-Academic-Statistics-in-the-UK-revised.pdf [Accessed: July 2019].

Nesta. (2014). '*Model Workers: How leading companies are recruiting and managing their data talent*'. London: Nesta. Available at: https://www.rss.org.uk/Images/PDF/influencing-change/rss-nesta-model-workers-data-talent-recruitment-2014.pdf [Accessed: July 2019].

Organisation for Economic Co-operation and Development. (2016). '*Research Ethics and New Forms of Data for Social and Economic Research*'. Paris: OECD Publishing. Available at: http://dx.doi.org/10.1787/5jln7vnpxs32-en [Accessed: July 2019].

Royal Statistical Society. (2016). '*Data Manifesto Briefing Note 1, Putting evidence at the heart of policy debate*'. London: Royal Statistical Society. Available at: http://www.rss.org.uk/Images/PDF/influencing-change/2016/data-manifesto1-putting-evidence-at-the-heart-of-the-policy-debate.pdf [Accessed: June 2019].

Royal Statistical Society. (2017). '*Data Manifesto Briefing Note 4, Public trust and pre-release access to statistics*'. London: Royal Statistical Society. Available at: http://www.rss.org.uk/Images/PDF/influencing-change/2017/data-manifesto4-public-trust-and-pre-release-access-to-statistics.pdf.pdf [Accessed: June 2019].

Royal Statistical Society. (2017). '*Data Manifesto Briefing Note 5, Increasing access to local data*'. London: Royal Statistical Society. Available at: http://www.rss.org.uk/Images/PDF/influencing-change/2017/Manifesto%20Briefing%20Note%205%20-%20Increasing-access-to-local-data-June-2017%20PDF.pdf [Accessed: June 2019].

Royal Statistical Society. (2016). '*Data Manifesto Briefing Note 10, Education and skills for the data economy*'. London: Royal Statistical Society. Available at: http://www.rss.org.uk/Images/PDF/influencing-change/2016/Manifesto-Briefing-Note-10-Education-skills-data-economy-FINAL-25-Nov-2016.pdf [Accessed: June 2019].

Royal Statistical Society. (2015). '*Data Manifesto Briefing Note 7, Investment in research, science and innovation*'. London: Royal Statistical Society. Available at: https://www.rss.org.uk/Images/PDF/influencing-change/2015/data-manifesto7-investment-in-research-science-and-innovation.pdf [Accessed: June 2019].

Royal Statistical Society. (2018). '*Royal Statistical Society response to CDEI consultation by the DCMS*'. London: Royal Statistical Society. Available at: http://www.rss.org.uk/Images/PDF/influencing-change/2018/RSS-CDEI-submission-5-Sept2018.pdf [Accessed: July 2019].

Royal Statistical Society. (2016). '*The Opportunities and Ethics of Big Data*'. London: Royal Statistical Society. Available at: https://www.rss.org.uk/Images/PDF/influencing-change/2016/rss-report-opps-and-ethics-of-big-data-feb-2016.pdf [Accessed: June 2019].

The Royal Society. (2016). *'Data analytics: the skills need in STEM'*. London: The Royal Society. Avaliable at: https://royalsociety.org/~/media/events/2016/11/data-science-workshop/data-analytics-conference-report-16112016.pdf [Accessed: July 2019].