

## Evidence to the Royal Society and British Academy on Data Governance

Please give details of the organisation you represent (if any), and which of the following sectors it falls into:

- Other, or cross-cutting sectors

The Royal Statistical Society (RSS) is a learned society and professional body, with more than 8,000 members in the UK and across the world. We are one of the world's leading organisations to promote the importance of statistics and data, and have done so since we were founded in 1834. As a charity, we advocate the key role of statistics and data in society, and we work to ensure that policy formulation and decision making are informed by evidence for the public good. We are pleased to respond to the Royal Society and British Academy's [call for evidence on data governance](#)<sup>1</sup>, as we would welcome more comprehensive discussion and open public leadership on this topic.

### A) Governance for data use: priorities, needs and opportunities.

A major priority for statistics at present is the integrity, and integration, of new data sources. We can highlight three separate drivers of this:

- 'Big data' is driving substantial changes in data availability. Statistics and statistical analysis is needed to make sense of the increasing variety, velocity and variability of data which are produced by the growth of digital and communications technology. As we presented in a paper to the European Commission<sup>2</sup>, statistical methods are fundamental to and at the heart of the advances that come from big data.
- In government statistics, and in quantitative social science research, there is a substantial drive not only to integrate some new 'big data' sources, but also to make better use of the administrative data that is already collected by government. Access to administrative data could make data collection more efficient, and also has potential to improve the quality and the timeliness of the research and statistics that are produced.
- Finally, and accompanying these developments, there has been growth in open data publishing, on the part of government but also in other data driven sectors. This is when

---

<sup>1</sup> Royal Society & British Academy (2016) *Data governance: Call for evidence* (PDF). Available from: <https://royalsociety.org/~media/policy/projects/data-governance/data-governance-call-for-evidence-september-2016.pdf> (accessed: November 2016)

<sup>2</sup> RSS (2014) Response to the European Commission on Mathematics and Digital Science (PDF). Available from [http://www.rss.org.uk/Images/PDF/influencing-change/RSS\\_response\\_to\\_the\\_European\\_Commission\\_on\\_Mathematics\\_and\\_Digital\\_Science.pdf](http://www.rss.org.uk/Images/PDF/influencing-change/RSS_response_to_the_European_Commission_on_Mathematics_and_Digital_Science.pdf) (accessed: November 2016)

data which is not confidential is openly published as a resource for anyone to analyse or use, as has been the case for example with open mapping data. We have highlighted open data in our [Data manifesto](#)<sup>3</sup> as an important spur of data-driven innovation and growth.

Accompanying these three drivers, we would like to emphasise three broad sector-wide requirements for the consideration of this review:

### **A1: We need to develop the capacity to use new data sources appropriately**

The growth of potential data sources seems to outstrip the capacity to curate and use them in practice.

- In the commercial sector, there are leaders and laggards in data access and use, and in many cases data is under-used. Recent research by Nesta concluded that ‘finding talent with the right domain knowledge, the right mix of skills (e.g. data scientists), experience, and business know-how to apply data in a commercial context is much harder than finding people with the right technical skills (including data manipulation and analysis)<sup>4</sup>. Domain knowledge in this case includes understanding the theories that explain relationships in the data, and knowing data sources and their limitations<sup>5</sup>.
- With regard to government statistics, the Independent Review of Economic Statistics by Sir Charles Bean ([Bean Review](#)) considered that new techniques need to be adopted by statistical agencies, such as web scraping, text mining and machine learning, and that this could form an important complement to their existing activities, for example to cross-check data, to fill in gaps, and to explore the significance of new economic phenomena before undertaking more systematic measurement<sup>6</sup>. The Bean Review concluded that statistical services need to be better equipped to use these sources, and should grow the capacity to clean, match and analyse very large data sets.

As adoption of big data and open data increases, there is a pressing need to understand the standards and integrity of the data for onward use. We need access to *metadata* which describes

---

<sup>3</sup> Royal Statistical Society (2016) *The Data Manifesto* (PDF). Available from: [http://www.rss.org.uk/Images/PDF/influencing-change/2016/RSS\\_Data%20Manifesto\\_2016\\_Online.pdf](http://www.rss.org.uk/Images/PDF/influencing-change/2016/RSS_Data%20Manifesto_2016_Online.pdf) (accessed: November 2016)

<sup>4</sup> P. 6 in Bakhshi, H. Mateos-Garcia, J. & Windsor, G. (2015) *Skills of the Datavores: Talent and the data revolution* (PDF). Available from: [http://www.nesta.org.uk/sites/default/files/skills\\_of\\_the\\_datavores.pdf](http://www.nesta.org.uk/sites/default/files/skills_of_the_datavores.pdf) (accessed: November 2016)

<sup>5</sup> P. 20 in Bakhshi, H. Mateos-Garcia, J. & Whitby, A. (2014) *Model workers: How leading companies are recruiting and managing their data talent* (PDF). Available from: [http://www.nesta.org.uk/sites/default/files/model\\_workers\\_web\\_2.pdf](http://www.nesta.org.uk/sites/default/files/model_workers_web_2.pdf) (accessed: November 2016)

<sup>6</sup> ‘Strategic Recommendation D: Make the most of existing and new data sources and the technologies for dealing with them’, pp. 11-12 in Bean, C. (2016) *The Independent Review of Economic Statistics* (PDF). Available from:

[https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/507081/2904936\\_Bean\\_Review\\_Web\\_Accessible.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/507081/2904936_Bean_Review_Web_Accessible.pdf) (accessed: November 2016)

key features of the data that are collected. The Bean Review found that access to the detailed *microdata* is also more necessary than before, as it is crucial to be able to ‘dig below the surface’ of any data source to establish its limitations.<sup>7</sup>

As we see the rise of algorithms being used in decision making (for example, in predictive policing, and also for job hiring and credit scoring), the Review should also consider whether there are methods that the public can use to hold algorithms to account. We are aware for example that crime data recorded by police is likely to be biased, as crimes that occur in locations that are already frequented by police are more likely to be recorded in a dataset<sup>8</sup>. Different communities can also be more or less willing to report crimes to police, resulting in a different source of bias. A machine learning algorithm trained on such data could guide policing in ways that are more biased about particular areas and particular ethnic groups than would otherwise be the case. Concerns about algorithmic transparency and accountability are growing in public profile – for example a White House report this year focused on the topic of Big Data and Civil Rights in the United States<sup>9</sup>, and Germany’s chancellor Angela Merkel has called for big internet platforms to make their algorithms more transparent<sup>10</sup>. More consideration is needed of whether and how governance systems can scrutinise algorithms’ fairness and accountability.

Governance should encourage sufficient transparency about data, and data standards, with a view to improving the usability of data and ensuring it is used appropriately. Whether there are appropriate mechanisms to do this needs consideration for the future of data governance.

## **A2: We need to strengthen the legal basis for data sharing**

Our *Data manifesto* recognises that data-driven innovation requires well-targeted and proportionate sharing of data for processing and analysis. Privacy safeguards need to be built into any sharing of personal data at the outset.

---

<sup>7</sup> ‘Strategic Recommendation E: Become better at understanding and interrogating data’, p. 12 in the Bean Review (PDF), Available from: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/507081/2904936\\_Bean\\_Review\\_Web\\_Accessible.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/507081/2904936_Bean_Review_Web_Accessible.pdf) (accessed: November 2016)

<sup>8</sup> Lum, K. & Isaac, W (2016) To predict and serve?, *Significance*, 13(5): pp. 14-19, Wiley Online Library. Available at: <http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2016.00960.x/full> (accessed: November 2016)

<sup>9</sup> Executive Office of the President (2016) *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights* (PDF), Washington: The White House. Available from: [https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016\\_0504\\_data\\_discrimination.pdf](https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf) (accessed: November 2016)

<sup>10</sup> Connolly, K. (2016) Angela Merkel: internet search engines are ‘distorting perception’, *Guardian*, 27 October 2016. Available at: <https://www.theguardian.com/world/2016/oct/27/angela-merkel-internet-search-engines-are-distorting-our-perception> (accessed: November 2016)

In government statistics, the ONS has been [assessing the potential for administrative data to produce the next Census](#)<sup>11</sup>, and the UK Statistics Authority has been [seeking to improve the legal basis for the government and businesses to share such sources](#)<sup>12</sup>, most recently in the development of statistical requirements in the UK Government's Digital Economy Bill. Research by academic and third party researchers also benefits from access to the necessary administrative data for public studies, and efforts are also ongoing in the Digital Economy Bill to [improve the legal basis for researchers](#) to access data<sup>13</sup>.

Legislation is an important reference point, but what is done with data is also affected by the culture of professions and organisations – the law needs to be understood if it is to have proportionate effects. The impact of legislation on governance and leadership needs to be considered.

### **A3: We need to maintain public participation and public trust**

Big data and data sharing poses possible risks to individuals' right to privacy. There are many fields of statistics and research (official statistics, in public health and medical research, in social science, and in market research) that begin by using personal data about or from individuals, based on informed consent. Analysts might integrate a variety of sources of such data to inform conclusions; they typically do so to produce aggregate statistics which are de-identified, and the safeguarding of personal identities is taken very seriously. Nonetheless, statistics are produced by analysis of individual-level or personal data, and members of the public are important participants in this endeavour. If they object to the way in which statistical collections are developed, or think that such collections lack integrity, collections can be placed in jeopardy. An example of this can be seen in the UK government's recent 'care.data' scheme, which would have linked data from hospital records to data from GP records to be used for research purposes. Communication to the public about the scheme is widely regarded to have been inadequate; there were substantial public fears about how widely, and for what purpose, the newly linked data would be shared and used; and follow-up of requests to opt out of the collection was substantially delayed<sup>14</sup>. Care.data was paused and officially discontinued, as a result of public concerns.

---

<sup>11</sup> ONS (n.d.) 'Administrative Data Census Project' (webpage), available at: <https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject> (accessed: November 2016)

<sup>12</sup> UK Statistics Authority (2016) *Delivering better statistics for better decisions: Why we need new legislation for better access to data* (PDF), available from: <https://www.statisticsauthority.gov.uk/wp-content/uploads/2016/03/Delivering-better-statistics-for-better-decisions-data-access-legislation-March-2016.pdf> (accessed: November 2016)

<sup>13</sup> Cuthill, V. 'Dreaming about legislative change: an opportunity for data-driven research that must be seized' (webpage), *ESRC Blog*, 26 September 2016. Available at: <https://blog.esrc.ac.uk/2016/09/26/dreaming-about-legislative-change-an-opportunity-for-data-driven-research-that-must-be-seized/#more-1392>

<sup>14</sup> Flanagan, O. (2014) 'Care.data: a year of bungling and confusion', *StatsLife*, 25 November 2014, Available at: <https://www.statslife.org.uk/features/1921-care-data-a-year-of-bungling-and-confusion> (Accessed: November 2016)

However, when the context, safeguards, and purpose of collection are clear and well-explained, public fears diminish. Engagement with members of the public through focus groups and deliberative research has shown qualified support for data to be used for scientific, social or statistical research, where this has an intended public benefit, and where there are sufficient privacy safeguards in place<sup>15</sup>. The willingness of the public to support research can be seen in well-established data-linkage studies that are supported by participants. To provide examples we abridge here two case studies of success factors for ongoing studies, drawn from [a report published by the Wellcome Trust](#)<sup>16</sup>:

- The Avon Longitudinal Study of Parents and Children (ALSPAC) is a large birth cohort study established in 1991 based in Bristol, England. ALSPAC have followed-up the health, well-being and development of multiple generations of study family members. ALSPAC has also linked data from its participants with a wide range of administrative data sources about them, including hospital records, care records, education records, and demographic data. This has taken place on a transparent basis whereby participants are informed of how the study intends to use their administrative records, and are offered a means to object and withdraw from this process if they so wish. The study has demonstrated high levels of support from participants, and arrangements for consent are coupled with 'safe haven' governance arrangements to show that ALSPAC is a reliable custodian of data.
- The SAIL (Secure Anonymised Information Linkage) databank established in Wales in 2006 has uploaded and linked over 2 billion records from multiple health and social care providers. The databank saw an increase in participation from 40% to 70% of primary care providers in Wales from 2014-2015. This appeared to be the result of a successful drive by a dedicated GP engagement team. In response to the 'data trust' issues that were emerging from the care.data scheme in England, SAIL also reviewed its communication policies to ensure sufficient communication with the public, and transparency in the form of more internal and external audits. The author of this case study says that increased participation from GPs resulted from communication about the benefits of linked data research.

---

<sup>15</sup> Economic and Social Research Council (2014) 'Public dialogues on using administrative data' (webpage), Available at: [www.esrc.ac.uk/public-engagement/public-dialogues/public-dialogues-on-using-administrative-data/](http://www.esrc.ac.uk/public-engagement/public-dialogues/public-dialogues-on-using-administrative-data/) (accessed: November 2016)

RSS (2014) 'RSS research finds data trust deficit ...' (webpage), *StatsLife*, 22 July 2014. Available at: <https://www.statslife.org.uk/news/1672-new-rss-research-finds-data-trust-deficit-with-lessons-for-policymakers> (accessed: November 2016)

RSS (2016) 'Public supports sharing health data for research...' (webpage) *StatsLife*, 14 March 2016. Available at: <https://www.statslife.org.uk/news/2721-public-supports-sharing-health-data-for-research-poll-says> (accessed: November 2016)

<sup>16</sup> Public Health Research Data Forum (2015) *Enabling Linkage to Maximise the Value of Public Health Research Data* (PDF). Available from: <https://wellcome.ac.uk/sites/default/files/enabling-data-linkage-to-maximise-value-of-public-health-research-data-phrdf-mar15.pdf> (accessed: November 2016)

Governance arrangements are key to act upon the intentions of the legislative framework for both data protection and data access, and to check privacy safeguards at the outset of new data sharing developments. In particular, we believe any governance framework must ensure attention to privacy concerns is balanced with public engagement on the benefits of data use.

## **B) The data governance landscape**

**What are the governance frameworks, processes and organisations that affect your work? What reviews of governance are you aware of or involved in previously or currently? Please detail examples of organisations and systems for the governance of data use.**

Data protection law, which has recently been reformed in Europe, forms a key reference point for governance. The content of European reforms has been a matter of interest to us, and alongside many other organisations we co-signed a series of statements by the Wellcome Trust, and welcomed that the reformed law maintains substantial provision for data to be shared for scientific and statistical research purposes<sup>17</sup>. The government Minister for Culture Media and Sport has made clear that the reformed regulations in Europe will take effect in the UK, as the General Data Protection Regulation (GDPR) is scheduled for adoption in 2018<sup>18</sup>, so adoption will not be affected by the outcome of the recent referendum.

Domestic law and arrangements are also important. This Review will be aware that there are specific statutory gateways for data sharing, and beyond this, non-statutory policies and codes of practice. The Law Commission completed a [scoping review of data sharing between public bodies](#) in 2014<sup>19</sup>, which found that approaches to data sharing were in some cases overly cautious, that statutory provisions for data sharing were in need of further review, and that the work of the Information Commissioner's Office and other bodies which provide advice and guidance should be further explored.

Regarding the law for official statistics, we have welcomed the Digital Economy Bill's intention to amend the [Statistics and Registration Services Act 2007](#), to strengthen the basis for government and business to share data for statistics. This amendment was called for by the Bean Review and has the support of the UK Statistics Authority, whose [briefing](#) addresses why they need new

---

<sup>17</sup> Thompson, B. (2016) Analysis: Research and the General Data Protection Regulation (PDF), Wellcome Trust. Available from: <https://wellcome.ac.uk/sites/default/files/new-data-protection-regulation-key-clauses-wellcome-jul16.pdf> (accessed: November 2016)

<sup>18</sup> Denham, E. (2016) How the ICO will be supporting the implementation of the GDPR, *Information Commissioner's Office blog*, 31 October 2016. Available at: <https://iconewsblog.wordpress.com/2016/10/31/how-the-ico-will-be-supporting-the-implementation-of-the-gdpr/> (accessed: November 2016).

<sup>19</sup> Law Commission (2014) *Data Sharing between Public Bodies: A Scoping Report* (PDF), available from: [http://www.lawcom.gov.uk/wp-content/uploads/2015/03/lc351\\_data-sharing.pdf](http://www.lawcom.gov.uk/wp-content/uploads/2015/03/lc351_data-sharing.pdf) (accessed: November 2016)

legislation for better access to data. Although we are in favour of the principle, [our evidence to the Bill's scrutiny committee](#) has questioned whether the clauses proposed are strong enough to have the intended effect<sup>20</sup>. Changes to the regulation of UK official statistics may also be needed; the [Bean Review](#) recommended that the UK Statistics Authority's regulatory function should be strengthened<sup>21</sup>.

Broadly, in the area of softer law and governance arrangements, we would like to highlight the conclusions of our workshop report [The Opportunities and Ethics of Big Data](#)<sup>22</sup>, which recognised a need for more proactive leadership on data ethics. A National Council exists in the United States, and we have called for a new national Council for Data Ethics to proactively consider the ethical challenges of new data science developments in the UK.

Finally, drawing on our experience of government data and statistics, we believe that the proliferation of data across government needs clearer governance, leadership and capacity across all work areas, both for analytical/statistical and service delivery purposes. It appears that the current mechanisms for data governance in government are too fractured, which risks a lack of strategic oversight for this key area of work. The way in which governance on key issues has been helped or hindered by the complexity of the UK's data landscape might be considered in this review.

RESPONSE ENDS

*Submitted by RSS' Policy and Research Manager, 4 November 2016*

---

<sup>20</sup> Royal Statistical Society (2016) *Briefing on the Digital Economy Bill 2016* (PDF), available from: <http://www.rss.org.uk/Images/PDF/influencing-change/2016/RSS-evidence-on-research-and-statistics-in-Digital-Economy-bill-Sept-2016.pdf> (accessed: November 2016)

<sup>21</sup> 'Strategic Recommendation F: Strengthen the governance framework so as to help support the production of high-quality economic statistics', pp.13-15 in Bean, C. (2016) *Independent Review of UK Economic Statistics* (PDF), available from: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/507081/2904936\\_Bean\\_Review\\_Web\\_Accessible.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/507081/2904936_Bean_Review_Web_Accessible.pdf) (accessed: November 2016)

<sup>22</sup> Royal Statistical Society (2016) *The Opportunities and Ethics of Big Data* (PDF), available from: <http://www.rss.org.uk/Images/PDF/influencing-change/2016/rss-report-ops-and-ethics-of-big-data-feb-2016.pdf> (accessed: November 2016)