ROYAL STATISTICAL SOCIETY

DATA | EVIDENCE | DECISIONS

RSS RESPONSE TO DCMS CONSULTATION ON 'DATA: A NEW DIRECTION'

19 November 2021

1. Introduction

The Royal Statistical Society (RSS) is an academic, professional and membership organisation for statisticians and data scientists. Rather than commenting on the consultation as a whole, we have identified a number of areas that are of particular interest to our members and we have made some recommendations in those areas.

2. Research access to data

A statutory definition of scientific research

It is not clear that it is necessary to create a statutory definition of scientific research to give greater certainty to researchers. The same objective could be achieved through strengthening and publicising guidance.

Furthermore, it is not clear that there is a robust enough definition of scientific research to serve as a statutory definition. Any option that we have considered seems either too permissive or too restrictive. For example, the definition proposed in the consultation – 'technological development and demonstration, fundamental research, applied research and privately funded research' – strikes us as both too permissive and too restrictive. It seems to consider any research whatsoever as scientific research, so long as it is privately funded. While at the same time it is not clear enough about its domain to set out whether study is restricted to the physical and natural world, or whether it also includes the social world.

From our perspective, it is important to be clear that social science research counts as scientific research in any definition. There is a growing demand in the social sciences for statistical and data analytical skills – and it is important that this growth is protected. In part this is because it will strengthen those disciplines, but it is also important because the students and early career researchers working in that environment will develop skills that can be applied once they graduate or leave academia. A statutory environment that stymied social science research risks disincentivising this move in these disciplines.

Lawful grounds for research

We welcome efforts to clarify lawful grounds for university research – enabling university ethics committees to more straightforwardly and precisely identify the lawful grounds for research would be a helpful step. Smaller institutions, in particular, are affected by a lack of clarity here – and this risks disadvantaging researchers in those organisations.

The RSS is not convinced that the best way to achieve this is to create a new, separate lawful ground for university research. There are two challenges here that concern us. First, it is important to consider public perception. It would be potentially damaging if the new lawful grounds could be portrayed as allowing essentially any university research – if the public feel that the safeguards that are in place on personal data use and reuse are not sufficiently robust, there is a risk that people will be less willing to consent to their data being used in the first place. And this, in turn, would mean that any analysis conducted would be much less useful as the data it was based on would be highly selective. On the other hand, if – on order to maintain public confidence – overly robust safeguards are introduced, then there is a risk that useful research is discouraged. Of course, non-university bodies, public and private, conduct data-driven research too, and much research is collaborative. Any new lawful ground would have need to recognise this; in which circumstances the concerns above remain relevant.

It is not at all clear to us what type of safeguards would be sufficiently reassuring to the public while also not discouraging potentially useful research. If the issue can be addressed through improving guidance to universities so that they can precisely identify the correct existing legal grounds, that would seem to be a preferable approach.

3. Consent and legitimate interest

Public Understanding

Our view is that part of the challenge in working within the UK's existing data governance framework is that public understanding of issues around consent and legitimate interest is lacking.[1] If the system is going to be improved, it is important that there is a commitment to improving public understanding of issues around consent. This is more than just a matter of improving the guidance and requires a strategy for public engagement to build an understanding of how personal data is used. There are examples of good practice in this – earlier this year the Geospatial Commission launched a public dialogue to start a conversation with the public around location data. Encouraging this type of work – for different use cases and data types – should be a central part of the government plan.

Part of the importance of this is that an engaged public will help businesses to think about which data they collect and why. We are concerned that in current business to business services, data is processed without a legitimate interest (eg, an email system that a business uses might include the time that an email was opened rather than just the fact that it was opened). Data is increasingly processed in this serviced way and improved public engagement is an important part of the way in which better practice can be driven.

4. Machine learning and AI

Automated decision-making and trustworthiness

If data-based technologies are to be accepted by the public, both the technology and the organisations using it will need to be trusted. Trust is not something that is automatically given by the public – in order to be trusted, organisations and systems must demonstrate trustworthiness. This is an important and influential idea – the UK Statistics Authority's Code of Practice for statistics has trustworthiness as its first pillar. Trustworthiness, in the context of the Code, "comes from the organisation that produces statistics and data being well led, well managed and open, and the people who work there being impartial and skilled in what they do".

The consultation document talks about the need to build trustworthy and fair AI systems – this is welcome. As is the recognition (p.37) that this depends upon the intelligibility of individual decisions taken. However, it is also important to emphasise that the organisations developing algorithms need to demonstrate trustworthiness.

In connection to these issues we would like to highlight a recent report by the Office for Statistics Regulation (OSR) which – in the wake of issues around the proposed use of an algorithm to award A-Level and GCSE grades – reviewed the use of algorithms that are intended to be applied to individuals. Their report, *Ensuring statistical models command public confidence*, is highly relevant to the topic of trustworthiness and the lessons that they draw for organisations developing algorithms are important (p.62). Here there is one that we wish to highlight:

[1] We are not aware of quantitative research that has been conducted into this area – but the Open Rights Group have conducted interview-based research *Public Understanding of GDPR* which suggests that there is a high level of awareness that data processing requires consent, but a low level of understanding of what consent means.

Organisations should "meet the need and provide public value: in this context it is particularly important to engage with affected groups to test and ensure the acceptability of any new approach."

We would recommend that any organisation working in this field pays close attention to the OSR's review. In connection with the government's plans to allow personal data to be used for building trustworthy AI systems without additional permissions, we think it is worth asking whether this comes at the cost of engagement and whether that might impact trust in the organisation developing the AI system.

Avoiding bias in the application of AI algorithms in decision-making is of critical importance, but it is important to acknowledge that just because an algorithm is unbiased does not make it accurate. For example, an automated process which makes decisions entirely at random can be unbiased. It should also be a requirement of developers of automated decision-making processes that they consider the accuracy of their algorithms and the consequent uncertainty associated with the decisions made.

An inclusive understanding of machine learning and AI

There is a considerable overlap between AI, machine learning and statistical modelling and prediction. The best example of this is regression analysis, a statistical technique used for modelling, prediction and decision-making, which is also used within machine learning and AI applications. Indeed, some AI applications are simply standard regression models. Conversely many applications of regression modelling for decision-making purposes would not be thought of by their developers as AI. Clearly, the legal framework cannot be based around how a developer chooses to describe their application. It must be based on robust and consistent considerations of the purpose of the activity and how individual data are used to achieve that purpose. We also note that the balance between human input and automation in decision-making algorithms is relevant both in model building and in the application of the model in decision making. The current proposals only seem to consider the latter.

5. Competition Duty

A role for the Office for Statistics Regulation

The consultation document includes a proposal for the creation of a Digital Regulation Cooperation Forum (DRCF). We would like to see some involvement of the OSR in the DRCF. Some of the issues that the DRCF will likely be looking at will involve statistics and where that happens, it is important that the statistics regulator is also involved. In October 2021 OSR published some guidance for models – setting out how the principles in the Code of Practice for Statistics can be applied to support good practice in the development of models. This is a good example of an area where the DRCF might be interested in conducting work and where the OSR has already been leading efforts.