The Campaign for Better Statistics response

to the UK National Data Strategy (NDS) consultation.

Preamble: This response to the consultation requested by the Department for Digital, Culture, Media & Sport has been prepared by the Campaign for Better Statistics. Originally titled the Campaign for Better Business Statistics, the Campaign grew out of dissatisfaction with various aspects of UK business statistics, as provided by the ONS, and the belief that they were failing to capture many of the details of the so-called Gig economy. Our continuing investigation of that belief and the manner in which the ONS responded both to the Bean report and more recently the Covid-19 pandemic, has led to a more general interest in the manner in which the UKSA exercises its responsibilities. Since statistics can only ever be as good as the data they depend upon, it follows that the Campaign also has a strong interest in the proposed National Data Strategy and we therefore provide our comments below. In doing so, we follow the format of the consultation, taking each question in order; we also make reference to the policy paper published on the 9th September, as the consultation was announced[1]. In the main we are strongly in favour of the proposed policy and welcome the further commitment to open government that underlies many of the proposed details.

Q1 Taken as a whole, do the missions and pillars of the National Data Strategy focus on the right priorities?

We ***somewhat disagree***. Although the mission of growth may be important we emphasise that it should be ***sustainable*** growth. We are also concerned that the benefits of data should be accessible to, and inclusive of, all citizens and these aspirations should be a specific objective of the overall mission. The description of the missions and pillars is therefore missing the foundational principle of Open-ness.

The executive summary of the policy paper mentions several categories of data - administrative, operational, transactional, analytical and statistical. We believe there should be a presumption of timely and unrestricted publication of all data collected by public institutions, and where data is redacted or aggregated this should be only for very limited reasons such as personal privacy and with explicit case-by-case justification. Decisions to withhold data should be subject to regular independent review.

We take this view because we believe it to be an implicit consequence of the executive summary which says: "*Data is a non-depletable resource in theory, but its use is limited by barriers to its access – such as when data is hoarded, when access rights are unclear or when organisations do not make good use of the data they already have. These barriers undermine the performance of public services and our economy, risking poorer outcomes for citizens.*"

The summary goes on to further strengthen this intention:

[1] https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy

"We will ensure that data can be leveraged to deliver new and innovative services, promote stronger competition, and better prices and choice for consumers and small businesses"

Innovation in commerce and society at large tends to be bottom-up, not top-down. So if the government is truly to promote innovation in the data sector it must make data available without barriers of cost or formality to small unfunded businesses, even private individuals. If "hoarding" of data is acknowledged as an issue then the government should lead by example and open up its own holdings of valuable data. When a public sector organisation does not make good use of the data it already has, then perhaps members of the public they serve can.

There is a latent danger in the "***Data Foundations***" pillar that the aspiration to create standardised data formats will only deliver data after a lengthy process of deliberation, with the pace of change exceeding the rate of formalisation. In our view, data users will readily accept data in arbitrary ad hoc formats so long as they get the data. Imperfect data today is better than perfect data someday. So it is vital that standardisation is not a prerequisite for publication. Data provision is a public good that allows for stakeholders to make better and quicker decisions. In this Covid-19 recovery post Brexit environment, data hoarding creates a barrier to both private and non-profit sector recovery.

Q2 - The COVID-19 pandemic has required radical and urgent innovation. This is exactly where small, informal, creative teams were most likely to generate a solution. For track-and-trace, for instance, investing one million pounds in each of a hundred or so alternative proposals was more likely to produce an effective result quickly than investing hundreds of millions in just one or two 'comprehensive' proposals. For example, a lot of track-and-trace data is implicitly available from e.g. contactless travel data like Oyster in London, credit/debit card transaction activity and contactless payments and mobile phone tracking. All of these link individuals to times and places and while not comprehensive could provide a very helpful aggregate picture of personal contacts. The obstacles to accessing this data were regulatory rather than technical and we believe public consent could have been obtained to the necessary disclosures, if the benefits had been properly explained.

We do not understand why there should be a separate process for Health and Social Care data. The only thing that makes these special compared to other data is that they are particularly important to people's wellbeing. So in our view all the benefits to be gained from improving utilisation apply even more strongly to health and social care data, far from being an exception to the present exercise, they should be one of its priorities. Thus we do not see the necessity for a separate "Data Strategy for Health and Social Care" all the concerns for data privacy remain the same whether in respect of health or other data.

Q3 - Is not GDPR there precisely to protect anyone affected by the equality act? Perhaps GDPR needs a more prominent role in the strategy, as a foundational principle or a pillar in itself. We are aware there is a conflict between the needs of GDPR and the needs of open-ness (akin to the present lockdown to protect public health versus re-opening the economy to protect public wealth). It could possibly be helpful if one of the explicit objectives of the strategy was to find the most effective synthesis from this antithesis, one that can evolve in pace with technological innovation.

Q4 - We feel that where statistical data are being collected it is important to have a nation-wide approach. The COVID-19 situation has created an impression of a disUnited Kingdom where each constituent country has chosen its own epidemiological metrics, and its own policy for time and

nature of publication. The Office of National Statistics in particular has sometimes seemed like the Office of English Statistics, sometimes the Office of England and Wales Statistics, and hardly ever like an Office of National Statistics.

When the nation as a whole needs data they should always be collected in a consistent manner and presented in the same detail for each constituent nation as well as in toto.

It is important to have comparable calculation processes for each part of the UK if the relative performance of the national and devolved governments is to be objectively comparable, to the benefit of all our citizens. It also seems inappropriate to develop differing standards and publication principles in each nation. We believe the UK Data Strategy should be UK-wide and that there should be buy-in from the devolved governments. It follows that when it may arise that other nations or multinational organisations have led (e.g. the US and EU) we should not be afraid to follow rather than re-invent.

We have more specific concerns about the Government Statistical Service and government sponsored surveys in general. We have made a more detailed response to the recent consultation ... but we believe that wherever survey data is used to support policy then the survey should have a strong statistical footing and that all methodologies, suppliers and algorithmic models used should be fully disclosed. Where methods are used that exclude certain groups (non-users of online services or "gig economy" workers), there should be evidence that these groups will not be disadvantaged by any ensuing decisions. When these groups are at risk of disadvantage then the use of fully representative research methods should be mandated, even if expensive. We believe this is crucial for genuine delivery of the Responsible Data pillar.

The Consultation document alludes to the Digital Economy Act 2017 and the Secure Research Service as operated by ONS. We argue that parts of the associated Code of Practice[2] from the 2017 Act (in particular Principle 4 "Public Interest" referenced in section 2.4) have a baleful effect on productive use of data and should be abandoned. The requirement to meet Public Interest mandates a gatekeeping process that necessarily slows the research process and eliminates the possibility of "blue sky" research, where there is no clarity of outcome at the outset. We believe the remaining aspects of the code, including review of outputs before circulation are sufficient to satisfy any public interest which has no relevance to confidentiality. Moreover we would argue that the process is inherently undemocratic.

In short we believe that the SRS as presently operated is only appropriate for the most detailed and most sensitive data and that the "Public Interest" is best served by openly releasing as much data as possible, consistent with the over-arching principles of GDPR. In which context we would draw attention to the processes used by the United States to ensure granulated data whilst preserving anonymity.

Q5 - We feel this question should be inverted. Not who stands to benefit, but which data sources have the most potential to add value? For instance, almost all the sectors mentioned, and others

[2] https://www.gov.uk/government/publications/digital-economy-act-2017-part-5-codes-of-practice/research-code-of-practice-and-accreditation-criteria

would benefit greatly from the full and free publication of Ordnance Survey mapping data. However, the detailed OS data is only expensively available and comes with re-publication restrictions. The priority seems to be generating a modest income stream to offset the cost of collecting and maintaining the data, rather than releasing the data and enabling a far greater value-add in the economy at large.

Q6 – Government has two primary roles as we see it:

Firstly, by extension of the argument in Q5 central government should be publishing far more data without charge or restriction. We mentioned the Ordnance Survey, but this applies also to the Met Office, hydrographical data, Land Registry data, local authority planning applications, register office data, county court judgements and many others. There should be a principle of timely, comprehensive and unrestricted publication of all public data by default. This is the single biggest step that central government can make to empower the data economy, and it is also cheap and rapid to deliver.

Secondly, the government should use its power as a bulk purchaser to insist that all data collected on behalf of the public should be delivered in non-proprietary formats that are fully disclosed and unencumbered by patent or other restrictions. This applies particularly to internet-of-things (IOT) data in sectors like health-care where e.g. data from wearable devices are typically collected in a private cloud in proprietary formats and thereby effectively appropriated by the manufacturer of the device. It should always be possible to direct the data stream to an alternative cloud or to rehouse the data from one cloud to another.

Unfortunately it is our opinion that the existing guidelines for open government are not being effectively adhered to in many cases, particularly by the Government Statistical Service.

Q6a - We believe these principles of disclosure of data and data formats should be pervasive of all sectors and applications. This is one area that government and stakeholder interests align.

Q7 - We strongly agree. We believe the insistence on open formats for data supply mentioned in our response to Q6 is one part of the government's role. We also reiterate our response to Q1 that while standardisation of data formats is important, publication of data in existing formats must be the priority. Development and implementation of formatting standards will take years and cost millions, releasing data in existing formats costs little and can be done almost immediately. With regard to other standards we feel that there should be legislation that actually punishes deliberate falsification of data, particularly by organisations. We believe that, to some extent, GDPR already represents an important precedent for such legislation. Moreover, governance could be positively impacted by extending the powers of the Information Commissioner's Office. We support the broad direction of this reform element.

Q8 - The main barrier that SMEs face in using data effectively is getting access to it. As well as implementing the principles of our response to Q6, the government should perhaps implement something similar to a Freedom of Information request for data (Freedom of Data request?), whereby any entity or individual can request access to data collected by a public body. The body would respond by identifying the relevant information and publishing it on a government website for the benefit of the requestor and any subsequent person with an interest in the same data.

Where the data are generated regularly, they would then be published regularly as they became available. Although this is a piecemeal approach and not apparently consistent with the Data Foundations pillar, it does have the advantage of being rapid and delivering data that people actually want to use. The experience of operating in this way can then inform the longer-term process of enhancing data quality and reusability.

Q9 - The government should, as part of the Data Skills pillar, be promoting awareness in the public of the value of their **private** data and encouraging them to take ownership of it. Holding personal data in the context of GDPR is an onerous task exacerbated in the health sector. The model can perhaps be inverted so that it is the individual and their (human and virtual) agents who become responsible for the custody and disclosure of their own data. An approach like SOLID[3] (developed by Sir Tim Berners-Lee) along with the opportunities from Smart Data can avoid the replication of personal data across many data controllers and the consequent risks. In short, consumers should be encouraged to both control and to exploit their own data. This is particularly important to offset the growing power of organisations such as Google, Facebook and Apple who should also be encouraged to make their data more widely available.

Q10/Q11/Q11a - Regulation can be the enemy of innovation especially in fast-moving sectors where it is easy to target yesterday's problems. Our view is to avoid intervention if at all possible. The government role is much more important when monopolies start to emerge. Government needs to avoid mistakes like creating purchasing policies that mandate the use of products rather than standards as happened with Microsoft Office in the 1990s.

We strongly agree with the objectives of the CDEI and therefore its importance, however from outside government it is hard to understand how management of data and data policy is shared between new and existing entities, some mentioned and some not mentioned. To name but a few:

- Information Commissioner
- National Statistician
- Government Statistical Service
- Centre for Data Ethics and Innovation
- Chief Data Officer
- Data Standards Agency

The Strategy should determine and explain what the role of each of these is, what authority they have under what Statutes, how conflicts between them are arbitrated, which are within the Civil Service and where, and which are QUANGOs? Maybe it should also explain how they are all necessary?

On balance, taking into account the needs of stakeholders, there is a need for an umbrella organisation with oversight of the data strategy and of National Statistics with the statutory authority to ensure privacy, ethical and quality standards are maintained whilst offering the

[3] https://en.wikipedia.org/wiki/Solid_(web_decentralization_project)

maximum of public access and accountability. It is readily seen that the individual lies at the heart of this requirement so we suggest that a suitable title might be the Centre for Public Engagement.

Q12 – **Quality, availability and access**: There already exist standards for each of these in respect of statistics and the requirement for open government. In our opinion these are not always observed
Standards and Assurance: there could certainly be improvements in this area. With raw data there need to be clear guidelines as to origin and timeliness as well as accurate description of content.
Capability, leadership and culture: yes, good examples of all these are preferred but can these be assured!
Accountability and productivity: Leadership should always be accountable but we are uncertain as to what is meant by productivity.
Ethics and public trust: We would place this at the top and re-emphasise that this is also vital to every aspect of the work done by the GSS.
In our opinion public engagement will be the key to success

Q13 – With regard to the **Data Standards Authority**, we note that this was only recently formed and there is little information available. It should be noted that our belief is that good standards emerge from practice and not the other way round. Data can be used, even without standards, so long as formats are disclosed. We are very concerned that the process of standardisation could delay the release of data that could be creating early value. It is crucial that the Data Standards Agency is an enabler of the effective use of data, rather than an obstacle to its use.

Q14/Q15/Q16/Q17 - We comment only with respect to Question 14. A vital component of data security is strong end-to-end encryption and yet certain government agencies insist that it should be compromised for the sake of access to certain communications. The entire digital nation, from domestic routers to the systems controlling critical infrastructure will be threatened if secure encryption is deliberately weakened. This threat is existential and dwarfs any benefit that may arise from being able to monitor some malicious actors and even then only if these actors are themselves negligent of their own security. So our response to Q14 is a negative one - providers should not be forced to open their systems to intrusion or they cannot guarantee any of data security, continuity or resilience of service supply.

Q18/Q19 – we have no comment.