



ROYAL
STATISTICAL
SOCIETY



THE
CENTRE
FOR
PUBLIC
DATA

INSTITUTE
FOR
GOVERNMENT

Getting data right

Perspectives on the UK National Data Strategy 2020



This is licensed under a Creative Commons
Attribution-ShareAlike 4.0 International license

Open Data Institute

Contents

Background	4
Key findings	5
Pillar 1 (Event 1) Foundations: ensuring data is fit for purpose	5
Pillar 2 (Event 2) Skills: data skills for a data-driven economy and data-rich lives	5
Pillar 3 (Event 3) Availability: ensuring data is appropriately accessible	6
Pillar 4 (Event 4) Responsibility: driving safe and trusted use of data	6
Pillar 1 (Event 1) Foundations: ensuring data is fit for purpose	8
Introduction	8
Opening remarks	8
Discussion	9
Fixing the plumbing	9
Investment, cost and value	9
Types of data	9
Institutions	10
Enforcing standards	10
Publishing data	10
Public attitudes	10
Challenges and solutions exercise	11
Breakout group A: How do we tackle the cultural and coordination barriers to good quality data?	11
Breakout group B: What needs to be done to resolve the long-running problems of legacy IT and broader data infrastructure? And how can we drive better data discoverability in government?	12
Breakout group C: What should a cross-departmental governance mechanism with the authority to enforce standards across government look like? How do we drive aligned governance structures across government?	13
General comments	15
Pillar 2 (Event 2) Skills: data skills for a data-driven economy and data-rich lives	16
Background	16
Setting out the challenge	16
Proposals for addressing these challenges	17
Pillar 3 (Event 3) Availability: ensuring data is appropriately accessible	19
Background	19
Open data	19
Shared models for deriving value from data	20
International data availability and collaboration	21
Debrief by government attendees	22
Pillar 4 (Event 4) Responsibility: driving safe and trusted use of data	23
Background	23
Realising transparency and strengthening accountability	23
Public sector registers	24
Additional points to note	25
Data regulation and innovation	25

Better, not more, data	26
What's missing or needs development to deliver responsible data?	27
Define clearer outcomes	27
Data is not responsible, people are	27
Public engagement	28
Value stance	28
Further topics discussed	29
Synthetic data	29
International data transfers	29
Digital IDs	29
Resources and examples mentioned during the event	30
About us	31
About the Ada Lovelace Institute	31
About the Centre for Public Data	32
About the Institute for Government	32
About the Open Data Institute	33
About the Royal Statistical Society	33

Background

The National Data Strategy was launched for consultation on 9 September 2020.

In response, the [Ada Lovelace Institute](#), the [Centre for Public Data](#), the [Institute for Government](#), the [Open Data Institute](#), and the [Royal Statistical Society](#) worked in partnership to convene a series of events focusing on developing recommendations for the implementation of the National Data Strategy for the public good. The aim of this series of events was to coordinate input, feedback and constructive recommendations for the UK government about how the goals of the National Data Strategy can best be implemented for the public good, in pursuit of aims of encouraging growth and innovation while also protecting and empowering people using the tools of ethics, good governance and regulation.

The **'Getting data right'** series analysed the four Pillars of the National Data Strategy, and convened conversations amongst experts and practitioners about how these themes could best be addressed and realised practically and sustainably.

Pillar 1 (Event 1) | Foundations: ensuring data is fit for purpose

Pillar 2 (Event 2) | Skills: data skills for a data-driven economy and data-rich lives

Pillar 3 (Event 3) | Availability: ensuring data is appropriately accessible

Pillar 4 (Event 4) | Responsibility: driving safe and trusted use of data

We are indebted to all the participants in these events, and in particular the invited speakers whose presentation insights and expertise have informed this joint event summary note:

- Seb Bacon (EBM Datalab, University of Oxford)
- Stephen Blackburn (Leeds City Council)
- Lynn Currie (ICO)
- Phil Earl (DCMS)
- Ben Goldacre (University of Oxford)
- Joshua Harris-Kirkwood (DCMS)
- Dr Kristina Irion (Institute for Information Law)
- Jesper Lund (IT Pol)
- Raegan MacDonald, (Mozilla)
- Rosalie Marshall (Data Standards Authority)
- Emma McCoy (Imperial College London)
- Ricky McGowan (RSS)
- Dr Sanjay Sharma (Brunel University)
- Mathias Vermeulen (AWO)
- Jim Weatherall (AstraZeneca)
- Sharon Witherspoon (Academy of Social Sciences)

This report is a summary of the discussion and does not necessarily reflect the views of the speakers.

Key findings

These are some of the key ideas and themes that emerged in the discussions.

Pillar 1 (Event 1) | Foundations: ensuring data is fit for purpose

This event was led by the Institute for Government.

- **A focus on data foundations is welcome: government needs to ‘fix the plumbing’ rather than being distracted by shiny technology.** The other pillars will matter little if data foundations are not properly built and supported. Although there is growing recognition of this inside government, there is still sometimes a lack of ministerial understanding of how much work needs to be done on the basics.
- **Fixing data foundations will require long-term investment and maintenance.** How much does not fixing the plumbing cost government and society? At the same time, while framing questions in financial terms and particularly ‘market failure’ has its uses (especially in talking to HM Treasury), it could obscure wider benefits and opportunities.
- **Government must properly engage the public.** Public sector organisations must be aware of the sensitivities in using data, especially citizens’ personal data. If government wants to make more use of citizens’ data, it must bring the public with it and be clear and honest about policy goals and trade-offs.

Pillar 2 (Event 2) | Skills: data skills for a data-driven economy and data-rich lives

This event was led by the Royal Statistical Society.

- **Data skills are valued by a wide range of employers across all sectors.** It is important not just to focus on the growing demand for workers with specialist data skills and to ensure that people from a wide range of backgrounds are given the opportunity to develop these skills. This might mean: reintroducing something like AS-Level mathematics to increase the number of people taking maths past GCSE level and training teachers in all subjects so that they are comfortable in teaching data skills that are relevant to their particular subject.
- **There is a need for greater collaboration between universities and industry to produce graduates with appropriate skills.** This could be done by encouraging joint appointments for people to work at the interface of industry and academia. Which would require research councils to recognise diverse outputs so that this career path is properly valued.
- **Career pathways for people with data skills could be improved in, eg, operational research.** To take the NHS as an example, this employs a large number of people with data skills to analyse things like waiting lists and where there is need for new services – but these roles are classified as admin/clerical and they lack career pathways and opportunities for advancement. These roles should be properly classified as scientific/technical and leadership training should be provided.
- **It is important to develop a professional framework for data scientists with shared codes of practice.** Part of this means professional accreditation for people working in this area to help establish professional standards for data

science. This is the objective behind the Royal Statistical Society's new Data Analyst accreditation.

Pillar 3 (Event 3) | Availability: ensuring data is appropriately accessible

This event was led by the Open Data Institute.

- The UK has a real opportunity to ensure that data is made more open and accessible, so that its benefits can be felt in every corner of society while ensuring people, communities and our environment are protected from harms. To achieve these opportunities, **the UK will need an appropriate and workable framework built on open standards and open data where possible; transparency; clear and honest communication; ethics; public engagement and understanding; and good governance.**
- **For data accessibility to function well, new methods of governance may be required.** An approach such as data institutions, which are based on independent governance and oversight, may prove to be necessary and welcome in order to develop trust between all parties.
- It should be recognised that **sharing data can have an impact at both the individual and corporate level, and can potentially exacerbate asymmetrical distributions of power or opportunities.**
- **The challenges and considerations relating to sharing data are, in many cases, human and cultural factors rather than technical.** People must be given the necessary skills to use data with confidence. An appreciation of the value of sharing data must be engendered at all levels of an organisation.
- It is vital that **data is considered as part of any government procurement contract** and that this covers how the data will be collected, stored and made appropriately accessible and interoperable, and that this is sufficiently funded for the long term.
- The **importance of developing local, bottom-up data sharing projects was emphasised.** The local level is often where the greatest societal benefit can be realised and, perhaps more importantly, where the benefits can be directly felt. It is critical that local authorities are empowered to act in the public interest and protect public benefits.
- The issue of public permissions must be addressed. People often share data freely with companies but not with government. **Government must be clear about the purpose of sharing data, explain the risks and then deliver on promises while demonstrating good practice.**

Pillar 4 (Event 4) | Responsibility: driving safe and trusted use of data

This event was led by the Ada Lovelace Institute.

- **A commitment to greater government transparency is welcomed but will need development to create meaningful understanding.** Any transparency mechanism should be designed with different user groups in mind (publics, regulators, researchers, for example), and the proposed national engagement campaign on the societal benefits of the use of data must explicitly build in

participatory structures for citizens to be involved in scrutiny and decision-making.

- **Transparency mechanisms must cover a full evaluation of the sociotechnical system around an algorithm**, including explicit articulation of values being propagated through the systems. Models to consider include Helsinki and Amsterdam's algorithmic registers.
- **However well designed, transparency mechanisms do not equal accountability**: commitments to strengthen accountability need to be built into the National Data Strategy more explicitly. Options discussed for strengthening accountability include expanded requirements for fuller public 'umbrella' impact assessment to include data protection, algorithmic, equalities and human rights impact assessments combined into a single process.
- **The National Data Strategy needs to get more specific on which values or ethics it is scaffolding its approach around**, including clarity as to why it is prioritising those values, and acknowledgement of trade-offs. To ensure ethics translates into practice, legislation could set precise public policy goals that all parties need to meet, for example to protect against discrimination, health misinformation or election manipulation, and encourage a shift from a 'Can I build it?' to a 'Should I build it?' approach.
- **The National Data Strategy should put more focus on securing the social benefits that the data protection legislation offers**. For example, data protection impact assessments (DPIAs) offer the possibility to work upstream from products brought to the market, and eliminate potential harms from the start. A risk-based approach to innovation and leveraging data protection by design and default allows us to secure the social benefits of technological developments and upstream interventions. Participants warned that the UK should not move away from data rights under the rule of law, in favour of codes or ethics or self-regulatory principles.
- **Regulators can, however, be further supported to ensure the processes, tools and templates needed to clarify responsibilities and equip organisations to innovate effectively and compliantly are robust**. Clear guidance that organisations can emulate and adapt with confidence is not consistently in place. Greater collaboration between regulators (which might require additional resourcing) could be improved, as well as developing sandboxes to go beyond working with individual businesses, one at a time.
- **The National Data Strategy could be an opportunity to shift the paradigm towards better data rather than more data**, by making long-term investments to tackle real research challenges around creating technologies that are leaner and more responsible: incorrect and bad-quality data is cumbersome and costly, reducing productivity and blocking development.

A detailed summary of each event is below.

Pillar 1 (Event 1) | Foundations: ensuring data is fit for purpose

Introduction

The [Institute for Government](#) roundtable focused on 'data foundations'. The National Data Strategy uses the term to mean data that 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.'

But, as the Strategy also says, at present:

'data is not consistently managed, used or shared in a way that facilitates informed decision-making or joint working across government and the wider public sector. Data remains undervalued and underexploited.'

This is true of government at all levels and different parts of the public sector, but it may also be true of other parts of society and the wider economy.

This roundtable considered what could be done about these problems, how to overcome the cultural and other barriers to better data, how to 'fix the plumbing' of data infrastructure, how to solve long-standing problems like legacy IT, and what good governance and enforcement of standards should look like (and indeed, whether these are the right questions and this the right starting point).

It opened with some opening remarks from public servants working on data foundations at different levels of government, followed by a wider discussion with attendees mainly drawn from government and civil society. This was followed three breakout groups, each considering particular challenges drawn from the National Data Strategy and attempting to find solutions:

- Breakout group A: How do we tackle the **cultural and coordination barriers** to good quality data?
- Breakout group B: What needs to be done to resolve the long-running problems of **legacy IT** and broader data infrastructure? And how can we drive better **data discoverability** in government?
- Breakout group C: What should a cross-departmental **governance** mechanism with the authority to enforce standards across government look like? How do we drive aligned governance structures across government?

Opening remarks

After a number of years of decreasing momentum, participants agreed that data has raced up the agenda in recent months. 'Data' was already something this government

was focused on (for example, the creation of a new data science unit at 10 Downing Street), but the coronavirus pandemic has made clear just how critical the effective use of data is. Some argued that the coronavirus response represented a high watermark in the use and sharing of data, while others were quick to point out the failures, including around test and trace and a lack of sharing with local authorities. All acknowledged, though, that the crisis had created a burning platform, giving a renewed sense of urgency to government's efforts to use data better.

What the framework National Data Strategy refers to as 'Data Foundations' lie at the heart of this. If government cannot get the basics right – tackling problems with access to data, data management and architecture, standards, capability, literacy and more – the other pillars of the strategy are redundant.

Discussion

Fixing the plumbing

Participants were keen to stress that there is still an enormous amount of work to do in this area. Some pointed to a lack of ministerial understanding of how much basic work remains to be done in departments, with ministers distracted by the idea of shiny new technologies and failing to recognise the urgent need to invest in getting the basics right first. There are indications that this is beginning to change, with a growing appreciation of the need to 'fix the plumbing', even at the highest levels of government and beyond the usual data and digital circles. But participants were still keen for the Strategy to push this line more. This requires honesty about the nature of investment (see below), and a need to focus on bringing in and using the right skills in government, and to think about long-term maintenance.

Investment, cost and value

There were concerns, too, about how the question of investment is framed. Some felt that not enough had been done to work out how much failing to get the 'data foundations' right costs government and society, on the grounds that having a better sense of this should help quantify the costs and opportunities and drive further and faster action. At the same time, there was pushback against the idea that more needed to be done to identify existing instances of 'market failure', where not capitalising on the full value of data was resulting in missed opportunities for growth, jobs, R&D and more. A number of attendees felt that while a 'market failure' framing might appeal to HM Treasury, it was too narrow, making it difficult to talk about areas where government could usefully create new markets for wider public benefit through the better use of data. They stressed that more thought needed to be given to what it will be important for a digital state to do in future, and how data can be used to help.

Types of data

Participants were also concerned that some data types risked being overlooked in any future work, given the current scope of the Strategy. The management of the trillions of words that government produces in documents and emails every year, for instance, is not currently a focus for government, despite its fundamental importance to effective knowledge management and collaboration: generalists think it is a problem for the data people, and the data people think it has little to do with them.

Not all types of data are the same and therefore cannot be treated the same. The use of data has to be situated in context – the communities involved, the purpose of the data, the trade-offs involved.

Institutions

While government can do data well, people felt that this was particularly true where there was an authority (such as the Met Office or Companies House) with responsibility for aggregating the data held by organisations and publishing it consistently. Attendees pointed out that the much-lauded publication of London public transport data by TfL, for example, was a relatively straightforward task since this only involved one organisation publishing to agreed standards, whereas making similar data for the north of England open would require coordinating between many local authorities that collect and process data in different ways. Currently, in many cases, it is unclear who is responsible for bringing together these data sets, and what authority they have to compel others to work with them.

The new Data Standards Authority is beginning to fill this gap, to an extent, by working to ensure a common understanding of how we describe, store, query, reference and transmit data in the public sector. This should help government and other parts of the public sector understand whether data is fit for purpose, reuse data, join up services and better analyse data for improved policy outcomes. This work is only just getting underway, with the DSA focusing currently on three priority categories (out of 15 to 18 they have identified). Setting those priorities is a challenge in itself, but there is much further to go, with participants highlighting future hurdles to overcome around applying standards to data held in legacy IT systems, and making standards consistent over many years to allow for longitudinal analysis, for instance.

Enforcing standards

Agreed standards will certainly help, but it is currently unclear how they will be enforced. Attendees observed that changes in the law may be required: it was felt that the failure of the GDS registers programme, for example, was partly attributable to the fact that departments had no legal obligation to maintain them. Even then, though, it can be difficult to make sure that departments meet their obligations – as we have seen with declining rates of departmental FoI response rates in recent years. Clearer lines of accountability within departments and other public sector organisations are likely to be necessary to ensure work to strengthen the data foundations is carried out, ideally with a government Chief Data Officer at the head.

Publishing data

There was some discussion of what being open even looks like for the public sector in 2020 and beyond. Publishing csv files is positive and might help join up work and lay some stronger data foundations across organisations, but it will be lost on most of the population. A number of attendees felt that as well as focusing on the internal management of data, and its publication where possible, more attention should be paid to the presentation of that data, making it more easily accessible for members of the communities it relates to.

Public attitudes

This raised the question of public attitudes more broadly. Recent examples (such as the sharing of test and trace data with police authorities) have highlighted the sensitivities of

sharing data between particular organisations at particular times, and for particular reasons. There was broad consensus around the urgent need to use data more effectively, with substantial investments required in data foundations to enable better data sharing and use now. But participants also stressed the need for public sector organisations to be mindful of these sensitivities. Government needs to act, but to act effectively it must bring the public with it. This includes clarity about policy goals and honesty about the trade-offs that might be involved – something some participants felt could be brought out more strongly by the National Data Strategy.

Challenges and solutions exercise

Breakout group A: How do we tackle the cultural and coordination barriers to good quality data?



[\[original image viewable here\]](#)

The breakout session on cultural and coordination barriers focused heavily on issues of understanding and clarity.

A lack of understanding of the legal context for collecting, processing and sharing data, for example, was seen as a key barrier to using it effectively. With officials wary of – and strongly incentivised to avoid – making mistakes and lacking a strong understanding of what is and is not legal under GDPR and other regulations, many are overly cautious when handling data.

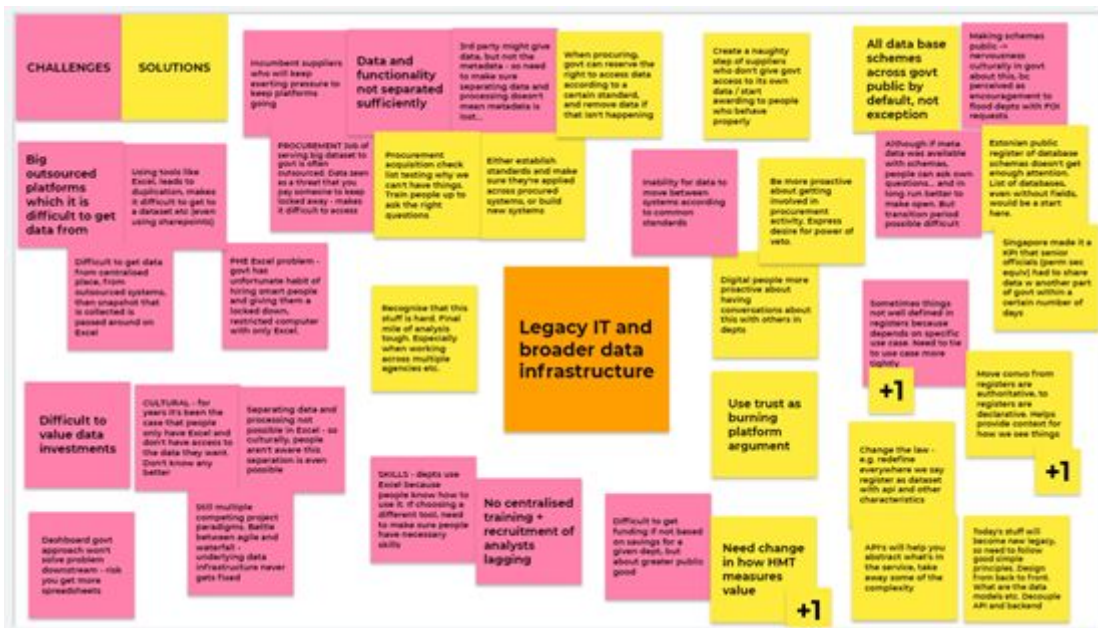
Attendees felt that this lack of understanding extended beyond the law, with responsibility for managing data often falling to non-technical officials who don't fully appreciate the potential power of data and understand the nuances of how to work with it effectively. Officials are not helped by poorly designed systems that make working with data a struggle, rather than supporting best practice (see Breakout group B for more). In other cases, data is seen solely as the responsibility of technical 'data' people, despite its relevance for work across government, leading to a disconnect between people who are collecting and processing data and those that are working with it. Worse still, in many

instances it is not even clear who owns a given dataset, and whose responsibility it is to maintain it.

A lack of understanding was also seen to characterise the attitudes of many senior officials and ministers. Participants felt that they too often see data as a 'background' task rather than a core element of effective government – although there are indications that this might be changing. They lack an appreciation of how difficult data can be to get right, looking for quick fixes for the symptoms rather than addressing the causes of problems with the use of data in government. This reflects a broader trend (with the framework National Data Strategy an example) of discussions about data in government being presented as 'win-wins' – all opportunity with no trade-off. While taking this line might help interest senior figures in data in the immediate term, attendees felt that it would ultimately prove counterproductive as the extent of the work and the difficulty of the decisions required to strengthen government's data foundations becomes apparent.

As well as greater honesty, there was a strong emphasis on the need for improved guidance to help overcome some of these issues. Translating DCMS advice on managing data into easy-to-follow processes for officials and frontline workers, for instance, would help those at 'the coal-face' grappling with questions of legality and unwieldy systems. A stronger overview of the cross-government data landscape and the lifecycle of key datasets would also help with assigning responsibility for data management, while offering an insight into ways of joining up the work of technical specialists and non-experts more effectively.

Breakout group B: What needs to be done to resolve the long-running problems of legacy IT and broader data infrastructure? And how can we drive better data discoverability in government?



[\[original image viewable here\]](#)

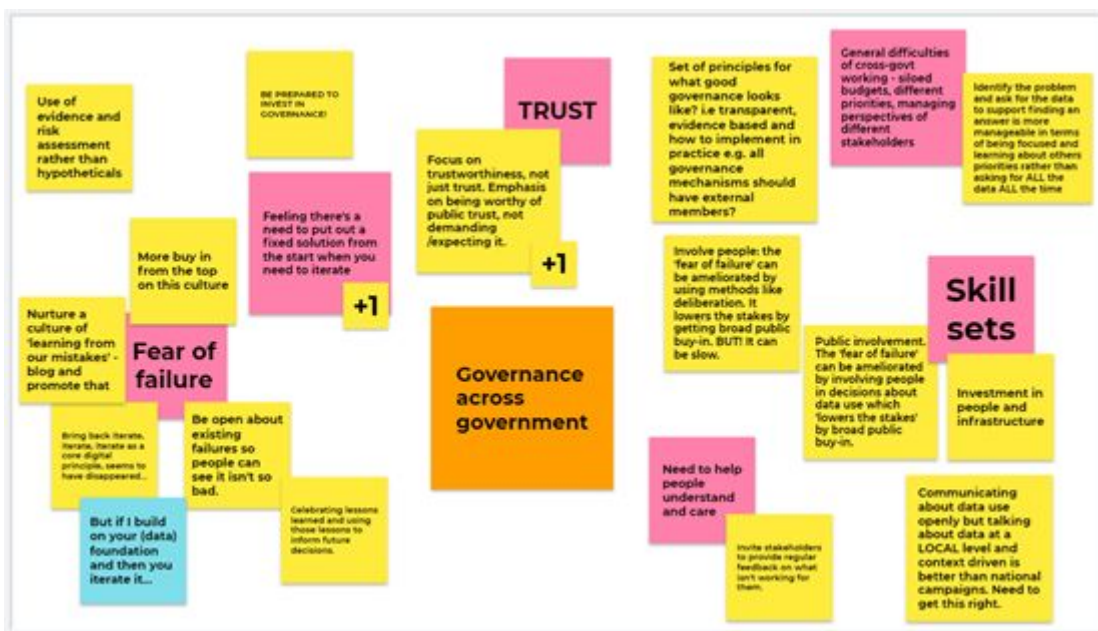
Government data is often stored in large, outsourced systems. The data in these systems can be difficult to access – attendees complained that while third parties might share the data on request, they sometimes refuse to share meta-data – and the data that is shared is formatted differently for each system, making it difficult to move data around government and join up datasets.

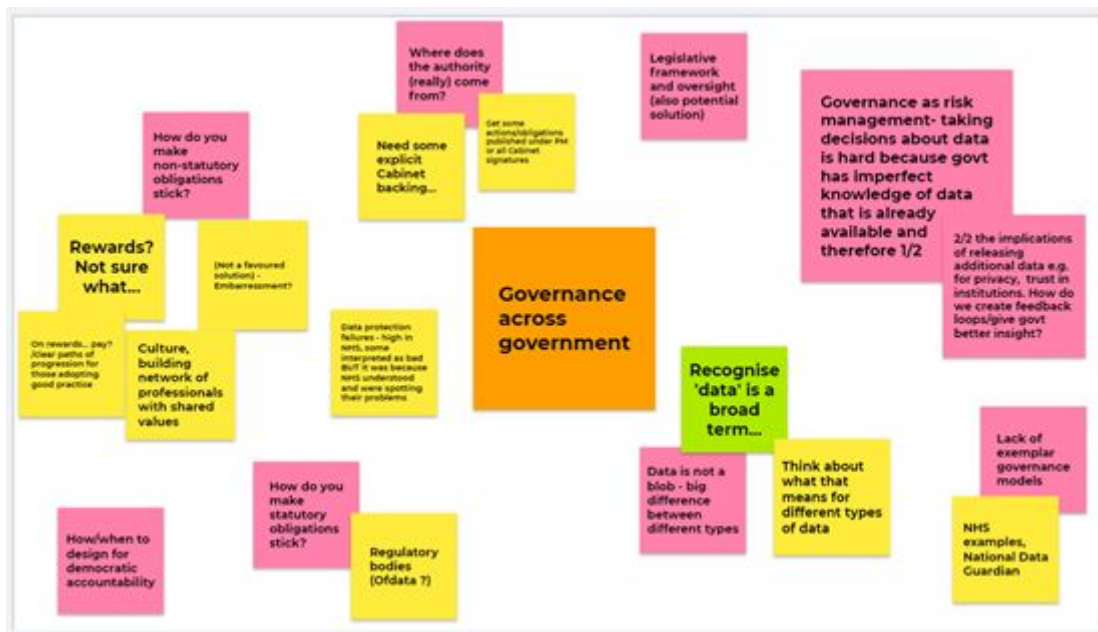
The data that is extracted is then typically put in an Excel spreadsheet to be emailed round, making effective processing difficult (there was some discussion of [the recent Public Health England Excel incident](#)) and leading to issues with information management as different versions circulate. This reliance on Excel is driven by a number of factors: departments use Excel by default because people know how to use it and using a different tool would require training; security concerns mean that officials are often only given 'locked down' computers with Excel and no access to other tools; and people have been using Excel for so long that they do not even know that better tools exist.

Participants had a number of recommendations to help address these issues. When it comes to procurement, they argued that government should reserve the right to access data stored in systems according to pre-agreed standards and remove the data if that is not happening. They felt that digital people should be more involved (and more willing to get involved) in procurement discussions, making sure that those making decisions properly understand what is required. A better training offer for those involved in procuring digital systems and for civil servants working with providers after procurement – as well as improved training to enable officials to move beyond Excel when processing data – was proposed.

They also argued for changes to improve data discoverability more broadly. As with Breakout group A, they felt that a better understanding of what data government holds, where and in what format, would be massively beneficial, suggesting that database schemas be made public by default. There was some emphasis on making better use of APIs and registers (making them less authoritative and more 'declarative', to help give context for how departments see things). And they stressed the need to be careful with the design of systems being built now – starting with the backend, ensuring robust data models are being used – since these will become the legacy systems of the future.

Breakout group C: What should a cross-departmental governance mechanism with the authority to enforce standards across government look like? How do we drive aligned governance structures across government?





[\[original image viewable here\]](#)

Much of the discussion around governance focused on a longstanding cultural barrier: the fear of failure in the civil service. Being open about existing failures and promoting a culture of 'learning from mistakes' would lessen the fear, but needed buy-in from the top of departments. Failures could, conversely, be a sign of some success – a department reporting breaches would show they could recognise them, while a department with none might be missing serious problems.

Another way of ameliorating the 'fear of failure' would be to involve the public and stakeholders throughout. Public deliberation could lower the stakes by ensuring broad buy-in, but this was likely to slow things down. There was also a discussion of trust in this context, and a focus on earning and not expecting trustworthiness (as explored by [the philosopher Onora O'Neill](#)). How to design for democratic accountability was also raised.

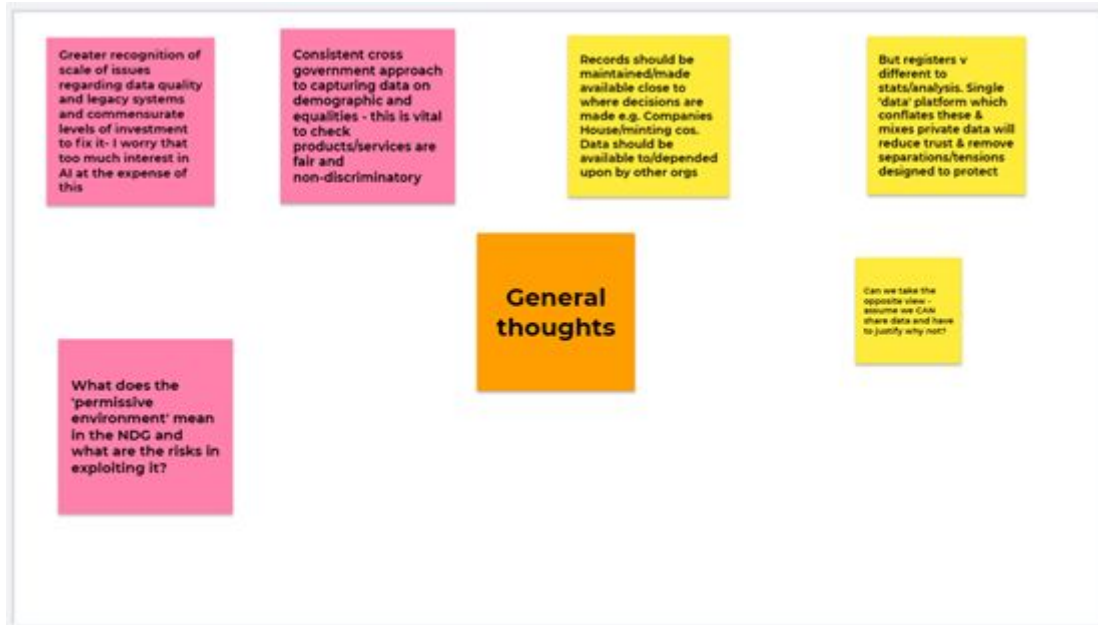
There was also discussion of where authority came from – the importance of getting Cabinet support and placing obligations on them was discussed – and how to make obligations stick, whether statutory or non-statutory. Most of this discussion focused on carrots – rewards, building culture and networks – rather than sticks – embarrassment, regulatory bodies.

The need to distinguish between different types of data ('data is not a blob') and situate discussion of it in context – at local level, rather than in national campaigns – was another theme. Additionally, thinking about exactly what data was required to solve a particular problem was more useful (especially in perennial problems of cross-government working and satisfying stakeholders) than asking for all the data, all the time.

A set of principles and good examples of data governance (including the work of the National Data Guardian) would be helpful. Such principles could include transparency and being based on evidence but also include more practical guidance, such as including external members as part of governance mechanisms.

As with other breakout groups, some basic steps (such as government knowing what data it held) would support better governance (for example, a better understanding of risk in taking decisions through knowing what data was already available).

General comments



[\[original image viewable here\]](#)

Pillar 2 (Event 2) | Skills: data skills for a data-driven economy and data-rich lives

Background

The National Data Strategy recognises that there is a growing demand for workers with specialist data skills. The Royal Society, in their 2019 report, [Dynamics of Data Science Skills](#), commissioned labour market analysis has shown that demand for specialist data skills had increased by 231% over the previous five years.

This is true across all sectors and industries: the importance of data skills is recognised by a wide variety of employers. Having these skills enriches understanding of individual, social and economic behaviours and is crucial in helping people to assess risk and make judgements. The current government has recognised this by indicating that it is placing a greater emphasis on making use of data analytical skills within the civil service.

There is also demand in the private sector for robust data scientists who can not just do the data engineering, but who can understand all the ethical ramifications of different kinds of data exploitation, and who can communicate the essentials of a complex data analysis to non-statistical experts – meaning that there is an appetite for employees with strong data skills from a wide range of backgrounds. It is important that data skills are not just confined to scientific and technical subjects: they are increasingly central to a wide range of careers – and it is positive that the National Data Strategy recognises that this means that data science must be taught across a wider range of subjects.

The [Royal Statistical Society](#)'s event focused on identifying some of the challenges involved in improving data skills and discussing what could be done in order to address these challenges.

Setting out the challenge

There are challenges across the skills pipeline – beginning in school and running through to the workplace – where there are specific challenges facing different types of employer. The discussion began by considering the situation in schools and universities, before moving on to discuss some specific challenges that face different types of employers.

Beginning at school level: statistics and data skills are not widely taught at this level. Ending AS Level maths as a stand-alone qualification reduced the number of people taking statistics modules and the core maths programme is only being offered by a minority of schools. Students who do not take mathematics at school do not get enough exposure to statistics and data skills. The problem is exacerbated by the fact that many school teachers, especially in non-scientific subjects, do not have the background to be able to teach the aspects of statistics that are relevant to their subject.

A similar pattern is observed at university level: data skills are increasingly important in non-STEM subjects, but they are not widely taught. As with schools, this is partly because teaching staff are not confident in doing so. At universities, there is also a

problem with retaining staff who do have data skills: they are prized by industry and it is hard for universities to hold on to them.

Discussion then moved to consider a couple of examples of how data skills are used in the workplace: operational research in the context of the NHS and in industry. The NHS employs a large number of people using analytical skills for operational research – things like modelling waiting lists and identifying what new services might be needed where. This profession is not properly valued within the NHS: these roles are classified as admin/clerical, progression is linked to line management and there are few opportunities for professional development. The shortage of opportunities for professional development is not helped by a lack of transparency when it comes to analytical methods and techniques. This both means that people don't learn from each other, but it also reduces organisational knowledge and efficiency.

In industry, data science skills are scarce and there is fierce competition for them. There is likewise a desire from the workforce to be able to upskill, re-skill or deepen their skills: some way for professional development in these areas to be formally recognised would be beneficial. Collaborations with academia are important – part of this means ensuring that universities are properly aligned with industry expectations as they develop new courses.

Proposals for addressing these challenges

The discussion also covered what could be done in order to tackle some of these challenges.

At school level, the reintroduction of something like AS-Level mathematics (a one-year, post-GCSE qualification) would be helpful as a way to introduce more students to statistics and data skills. In the past many students who would not wish to do a full A-Level in mathematics would take an AS-Level in the subject. The loss of AS-Level mathematics reduced the number of people who were learning statistical and data skills at school and there may be a case for reintroducing a shorter mathematics course after GCSEs.

As is recognised in the National Data Strategy, the teaching of statistics and other data skills must be widened out beyond mathematics and science subjects. to ensure that as many students as possible have some exposure to it. This cannot be effective without investing in upskilling teachers of all subjects, so that they are more confident in teaching students about the statistical/data aspects of their own subjects.

The situation is similar at a university level – statistics and data handling should be taught outside of STEM subjects. Government could incentivise universities to teach data skills in the widest possible array of subjects and train academic staff to properly teach students in their subjects.

In order to foster understanding between industry and academia – and ensure that universities are producing graduates with the skills needed by industry – is to encourage joint appointments for people to work at the interface of industry and academia. Research councils should recognise diverse outputs so that this career path is properly valued.

In the health service, leadership training for statisticians/data scientists should be provided so they can rise to senior levels in civil service. This process could be helped by

introducing a government Health Analytics Service (analogous to the government Statistical Service or government Legal Profession), with a head of profession, training programmes and ways of sharing best practice.

In government in general, where data analysts are classified in roles as admin/clerical they should be reclassified as scientific/technical and a national competency framework with pay scale and job descriptions should be introduced for this career path. And in operational analytics more broadly, we should aim to break open closed ways of working: currently results of analyses are routinely shared, but methods aren't. Opening this up will drive up quality and reproducibility and, in addition, sharing methods as routine is a powerful way of on-boarding people.

It is also important to develop a professional framework for data scientists with shared codes of practice. As noted in the National Data Strategy, the Royal Statistical Society has a role here to work with employers and universities to identify the skills needed for data scientists and accredit courses so that students and professionals can be confident in their quality. This work, as outlined in the meeting, is underway with the new Data Analyst accreditation. RSS can act as neutral broker between industry, academia and government to ensure that a framework here works properly from all perspectives.

Pillar 3 (Event 3) | Availability: ensuring data is appropriately accessible

In this event, led by the [Open Data Institute](#), we explored opportunities for developing or implementing some of the ideas, proposals, and actions identified in [Pillar 3](#) of the UK National Data Strategy in a way that is workable, sustainable, and ethical.

Background

Data availability covers several topics related to the collection and sharing of different kinds of data. Some well-known barriers to this include risk aversion, licensing, data hoarding, formatting, privacy and security, and equity. The aim of this roundtable was to discuss how to overcome these barriers, the role of government in this, and how to realise the full value of data in the economy.

The UK National Data Strategy defines ‘data availability’ as ‘an environment which facilitates appropriate data access, mobility and re-use both across and between the private, third and public sectors in order to generate maximal economic and/or societal benefit for the UK.’ Data availability is the third Pillar of the National Data Strategy, which outlines ambitions for improving data availability for the economy and society and for government, and international data availability.

The UK has a real opportunity with a National Data Strategy to ensure that data is made more open and accessible, so that value can be derived from it from every corner of society (public, private, the third sector or individuals and groups) while ensuring people, communities and our environment are protected from harms.

To achieve these opportunities, the UK will need an appropriate and workable framework built on open standards and open data where possible, transparency, clear and honest communication, ethics, public engagement and understanding, and good governance. Furthermore by seeking adequacy with Europe on the GDPR, the UK will be in an enviable position of getting to build and develop stronger data protections and data rights on the immovable foundations of GDPR. Such an opportunity will enable UK citizens and businesses to think dynamically about critical and meaningful areas such as data portability, interoperability and what consent will look like in the world of 5G.

For accessibility to function well new methods of governance may be required. An approach such as data institutions, which are based on independent governance and oversight, may prove to be necessary and welcome in order to develop trust between all parties: those providing data, those holding and sharing data and those using data.

Open data

The first session of the meeting focused on open data - the backbone of a modern, ‘smart’ digital economy. However, it should be recognised that data covers a spectrum from open to shared to closed. Understanding where on this spectrum any dataset falls and the conditions of access is a vital part of an effective open data ecosystem. There are a number of critical datasets such as those identified by the EU – geospatial, mobility,

demographic and meteorological. The UK government should similarly identify and make available such critical datasets and encourage others to do likewise.

It should be recognised that sharing data can have an impact at both the individual and corporate level. For example, it can exacerbate asymmetrical distributions of power or opportunities between companies. There's also a risk that the benefits of sharing data may not be realised equitably by those allowing access to their data. In addition, concerns about potential privacy and security breaches can make people or organisations wary of sharing data.

To overcome these barriers, it is essential to build trust. Building on the basic data foundations is also fundamental, and so standards will be important to ensure the quality and interoperability of the data. Audit trails and assurances will help, and trusted brokers or data trusts are an emerging option. Developing a well-rounded set of case studies for sharing data will be essential. These should articulate the added value of sharing data and demonstrate ways to mitigate risks, including how to build secure systems such as those designed by the ONS.

In general, it must be recognised that the challenges and considerations relating to sharing data are, in many cases, human and cultural factors rather than technical. People must be given the necessary skills to use data with confidence. This will require sufficient resourcing and the recognition of data skills as an asset. An appreciation of the value of sharing data must be engendered at all levels of an organisation. Often data sharing is blocked at higher management levels or director level because of the perceived risks and/or insufficient awareness of the potential benefits.

Cultural differences that exist between public, private or third sector organisations can be a source of friction in cross-sector data sharing initiatives. This can result from variations in technical vocabulary or common practices that differ between sectors. Even within an organisation, poor communication can lead to inefficient use of overlapping datasets while other datasets remain inaccessible and locked away. This sort of inefficiency is particularly an issue with cross-departmental data in government. To address it will require dialogues with multiple stakeholders in order to identify and categorise the data held and to agree on common approaches, legal frameworks and standards.

Shared models for deriving value from data

This session focused on potential solutions to encourage and enable the sharing of data, equitable returns of value, and making data discoverable and interoperable. Of particular interest were new models of data institutions such as data co-ops, data trusts and data unions.

Considerations were raised around government procurement projects and research projects. It was seen as important that any data generated as part of a project was treated as an asset, rather than as underlying physical infrastructure, and that there would be continued funding to maintain useful datasets beyond the time frame of the project. In general, it is vital that data is considered as part of any procurement contract and that this covers how the data will be collected, stored and made appropriately accessible and interoperable, and that this is sufficiently funded for the long term. [Building Information Modelling \(BIM\)](#) in the construction sector was cited as an example of this being done successfully.

The role of data trusts was discussed in relation to how transparency around data use can be improved and responsibilities for explaining to the public what their data will be used for, the benefits it might deliver and potential risks. The [US Data Collaboratives](#) were cited as a successful example. These are a 'new form of collaboration, beyond the public-private partnership model, in which participants from different sectors — in particular companies - exchange their data to create public value'.

The importance of developing local, bottom-up data sharing projects was emphasised. The local level is often where the greatest societal benefit can be realised and, perhaps more importantly, where the benefits can be directly felt. [Camden's open data platform](#) and the [London Data Commission](#) were cited as examples of this. Considerations around digitalisation affecting power structures was reiterated. In this regard, it is critical that local authorities are empowered to act in the public interest and protect public benefits. For this to happen, it is again vital that the public sector builds up the necessary human capabilities.

Global issues can be equally as fruitful as local ones. This is particularly the case within the research community and their efforts to tackle global challenges such as climate change. Recently, the Covid-19 crisis has shown the value of sharing data internationally. In particular, the [SAIL Database](#) in Wales was noted as a key national and international resource in fighting Covid-19. In general, it was stressed that the key to successful data sharing in research projects was to focus on the problem to be solved rather than the technology. Doing so will, in turn, drive considerations of trust and ethics and build the knowledge base. It will also promote a more holistic approach to data availability and wider stakeholder engagement.

International data availability and collaboration

The final session of the meeting focused on opportunities for international data sharing and collaboration. In particular, which international level problems could benefit from data sharing, what data is needed and what will be the impact of Brexit for trade and international relations.

The [Scientific Committee on Antarctic Research \(SCAR\)](#) was noted as a successful and long running model of data collection and sharing that has promoted peace and collaboration. The [Global Open Data for Agriculture and Nutrition \(GODAN\)](#) initiative was also noted positively. Oceanic, meteorological, and population datasets were cited as examples of critical national assets. However, it was observed that commercial opportunities that arise from such datasets rarely goes beyond the agencies that own the data. In the UK, health data and the ONS were seen as particularly strong assets. In the commercial sector, the nuclear and aerospace sectors were identified as good exemplars of data sharing and collaboration.

In general, it was recognised that global initiatives can be difficult. Different cultures can introduce barriers, and the need to equitably share value can be a challenge. It is therefore essential for any potential data sharing project to understand and protect national interests (particularly against multinational corporate interests), appreciate that local technical and human capacities may vary (such as the level of technical advancement of local health systems in relation to Covid-19), establish codes of practice, and develop trust. It was agreed that the UK had a leadership role to play while being sensitive to local interests.

In relation to the National Data Strategy, it is vital that the strategy has longevity. For this to be the case the government must clear about its intent in relation to GDPR and data adequacy post-Brexit.

Debrief by government attendees

The roundtable concluded with a debrief session for government attendees to reflect on the discussions. Some of the observations that emerged from those reflections are as follows:

The National Data Strategy must clarify what is meant by open data, the future of GDPR, how to improve discoverability, and how to build trust. A possible role for the National Audit Office in scrutinising government's use of data was raised.

The issue of public permissions must be addressed. People often share data freely with companies but not with government. Government must be clear about the purpose of sharing data, explain the risks and then deliver on promises while demonstrating good practice (cf ODI report [About Data About Us](#)). Public engagement and meaningful dialogue with stakeholders is vital. Issues relating to transparency, accountability and real choices for consumers are all challenging.

It is important to identify pinch-points and market failures that can be addressed by government, and that are unlikely to be addressed by others or naturally resolved over time; and also to consider what kinds of interventions and policy levers are within government's power - particularly given resource limits. It is vital to identify 'quick wins', while also recognising that some initiatives turn out to be more complex and challenging than anticipated.

The importance of purpose driven issues was highlighted. These could be fundamental, cross-cutting issues or more specific government priorities, and they cover areas such as climate change, jobs, and trade. There is a desire to do more to support SMEs.

Legislation is in place that is not always well understood, and some government departments continue to operate to older procedures. Culture change is important, but difficult in practice and adjusting to new processes and systems takes time. In this regard, Ministerial leadership is particularly important.

Pillar 4 (Event 4) | Responsibility: driving safe and trusted use of data

Background

[The Ada Lovelace Institute](#) organised a policy roundtable to stimulate discussion on the fourth pillar of the National Data Strategy on [Responsibility: Driving Safe and Trusted Use of Data](#).

The aim was to discuss what's missing from the strategy and come up with practical recommendations for the implementation phase, to ensure that data is used responsibly and in the interests of people and society.

The two main sections of the event were:

1. how to develop appropriate transparency mechanisms in the public sector for algorithmic systems
2. how to address actual or perceived tensions between innovation and data protection.

We also wanted to go beyond the specifics of the National Data Strategy and:

1. ask what the priorities should be
2. outline what we want to see the government leading on
3. offer challenges to the framings around 'responsible data' as outlined in the strategy.

Our main takeaways from the event, which build on the key messages from the discussion, are summarised below. We are indebted to the contributions from speakers and participants.

Realising transparency and strengthening accountability

A step to deliver responsible data in the National Data Strategy is to 'commit to addressing the need to develop appropriate mechanisms for increasing transparency and accountability in decisions made or supported by algorithmic systems, and for monitoring their impact'.

The broader commitment to greater transparency was welcomed, however it was noted that there was less specific content on accountability. Participants argued that transparency mechanisms should be explicitly seen as a means of achieving accountability for decisions made or supported by algorithmic systems.

Discussion cautioned against the idea that transparency might inherently enable accountability and articulated **the need to consider the sociotechnical system**: the problem is often not 'the algorithm' but how the algorithm is being deployed, what the goals are and who has the power to adjust the algorithmic system.

Furthermore, some argued that true transparency and accountability require an articulation of the values that are deemed important, and mechanisms to hold public and private actors accountable should ensure the systems meet those standards.

The National Data Strategy places important emphasis on the need to ensure the use of data is trusted. But **to be trusted, data use must be trustworthy**. Mechanisms for transparent, accountable, responsible data use discussed at the workshop will be successful if seen as actionable approaches towards developing more trustworthy uses of data.

Public sector registers

For enabling accountability in practice, one concrete recommendation is the use of **algorithmic registers**. Here we can learn from the cities of Amsterdam and Helsinki, and the work of [Saidot](#). As pioneers in this field, these cities have launched algorithmic registries to detail how city governments use algorithms to deliver public services, down to the level of the data sets used to train models and descriptions of how an algorithm is used.

In Amsterdam the registry is indirectly linked to **new standard clauses in procurement contracts that impose a duty of cooperation on the vendor**, to provide the municipality with all the information that may be required in order to explain how an algorithmic system works.

One suggested way to act on the importance of responsible data in practice was to **strengthen and require more widespread use of data protection impact assessments (DPIAs)** for converting the principles of law into reality. Different from privacy impact assessments, DPIAs are directed to all rights and freedoms of individuals and should go beyond data rights to check that executives are acting within their powers under the law. Performing an assessment of lawfulness and considering the system as a whole could be steps forward for ensuring more transparency and accountability for algorithmic systems.

Different assessment components, such as **data protection, algorithmic equalities and human rights impact assessments, could be combined into one process** under the umbrella of DPIAs, as prescribed by the General Data Protection Regulation (GDPR). To make this effective further guidance and standards would be needed, developing from the Information Commissioner's Office (ICO) existing guidance. Ireland's DPIA on the contact tracing app was suggested as an example of good practice.

This type of enhanced DPIA could be developed to be a requirement and be openly published. The process could evolve to include a public notice, a method of consultation and identification of certain data points to be gathered to assess efficacy and impact. This would enable people to understand how an algorithm was assessed and what the anticipated operation of it was.

Currently a key block to accountability is the absence of any feedback loop in the automated decision-making process. Further and ongoing analysis to check whether implemented measures worked is not performed. Therefore, data should be published to allow for ongoing monitoring of a system to determine whether it is producing just and lawful outcomes.

Additional points to note

- Any mechanism for transparency should be designed to ensure the recipients of information are able to use it. The recommendation was made that the government undertakes deep, deliberative and participatory consultation when it comes to designing what transparency is for. There should be explicit mechanisms for putting those who may be affected or marginalised at the centre of the process, noting the point that minimising bias in a technical system may be an impossible task if the entire service or system is built on discriminatory practices or thinking.
- A linked challenge was that the frame of transparency, accountability and fairness was currently limited to examining algorithmic systems in isolation. Should we expand our thinking to consider how we should deploy digital technologies to dismantle systemic racism for example, in other words – how do we establish fairness in an unjust society? Even expanding examining of algorithms beyond the technical to the sociotechnical doesn't account for the fact that the assemblage is embedded in racialised structures that are historically determined. One example provocation is that the school-to-prison pipeline can't be fixed by fairer or transparent algorithms. In effect the data-driven technologies legitimise and compound an already unjust system. Taking this frame should push us to ask some more challenging questions:
 - How can algorithms shift power?
 - What is the data generated for and in whose interest?
 - What values are encoded in algorithmic systems?
 - How can the pursuit of efficiency and prediction support the dismantling of systemic racism or other systemic inequalities?
- Roundtable participants welcomed 'lawful' as part of the definition of responsible data, but more thinking is needed on how that is implemented in practice and who would need to be involved – for example the government Legal Profession?
- Transparency mechanisms should contain information about how decision-makers will be trained to use systems (see Upturn report on information on meaningful information for scrutiny - link below).
- There was discussion around what point transparency should kick in – ideally the development pipeline should be made public and feedback sought so development can explicitly consider risks raised by the community. This would require courage and capacity so needs further thought, perhaps for high-risk applications like social care decision-making or asylum applications.

Data regulation and innovation

The National Data Strategy's underlying messages are around reaping 'the benefits of greater data use' and maintaining 'a fit-for-purpose legal and regulatory regime' while emphasising that the UK's 'data regime will support vibrant competition and innovation, building trust and maintaining high data protection standards without creating unnecessary barriers to data use'.

Data protection legislation can be perceived to be complex and difficult to work with. There were questions around what burdens are linked to perception, lack of knowledge and understanding, versus what are actual barriers in the law. Similarly, there has been discussion around approaches such as checking for compliance early in the innovation

process and providing more clarification, rather than a need to 'remove barriers' or any shift to a more permissive regime.

One recommendation was for the National Data Strategy to proactively incentivise accountability and responsible behaviours by using nudge behaviour economics theory to drive best market behaviours. In terms of actual barriers, it was pointed out that clear guidance that organisations can emulate and adapt with confidence is still missing. Organisations need practical guidance on how to develop strategies that encourage innovation which aligns with good governance. That will require leadership to drive good practice and shared approaches to facilitate wider culture change. We will need standard-setters such as the ICO to help organisations translate the National Data Strategy into an implementable process.

One message heard during the workshop is that DPIAs reverse the 'wait and see and intervene if there's harm' approach with a more proactive, social benefit first approach. Instead of waiting for innovators to bring new products and services to the market, only to find out that they are not compliant and enforcement is needed, DPIAs offer the possibility to work upstream and, if done right, eliminate potential harms from the start. This part of the legislation implies and indicates that we should be working to secure the social benefits that the data protection legislation offers.

A risk-based approach to innovation and leveraging data protection by design and default allows us to secure the social benefits of technological developments and upstream interventions. Weighing in risk – understood in a broad sense to include, for example, human rights and equalities assessments – and answering impact questions are facilitators in the process of innovation. In other words, this allows companies not only to innovate effectively and compliantly, but also to understand the legislation, which in turn upskills them to innovate more effectively and efficiently. The views put forward during the roundtable showed that compliance does not limit innovation, but instead is a competitive advantage as good data governance helps make better decisions and better products. For SMEs in particular, innovation and data protection should support and reinforce one another.

However, there are questions around how to streamline regulatory approval for businesses, something that is more challenging to do in the context of the GDPR, and whether this is a potential gap. Collaboration among regulators across sectors could be further improved, as joint interventions can create greater and more widespread impact than working with individual businesses one at a time. Some of the participants also expressed support for regulatory sandboxes and involvement from multiple regulators, together with the ICO.

There was also a clear warning voiced by some participants that the data rights that we currently enjoy under the rule of law cannot be sidelined in favour of codes of ethics, self-regulatory principles or by a 'writing your own rules and marking your own homework' approach.

Better, not more, data

Governments and companies are becoming more focused on data-driven decision-making. However, the message heard from this session is that better, not more, data collection leads to more innovation.

Incorrect and bad-quality data is cumbersome and costly, reducing productivity and blocking development. Instead, better data creates better models necessary for assisting in the decision-making process. It was argued that success for SMEs is supported by the quality of data, not the quantity. One way that companies could achieve quality of data is by instilling data protection and governance-first practices into their core business model, culture and brand.

The UK government has an opportunity to change the paradigm from more data is better. Incentives should be centred around what data you actually need and front-loading innovative processes that requires less data. The government has the opportunity to break the implacable 'move fast and break things' philosophy and think of long-term investments.

Organisations should not be encouraged to hoard data in order to innovate later. On the contrary, if we want better and more sustainable outcomes, the reverse approach yields more responsible end results. The recommendation was that the strategy should focus on the real research challenges around creating technologies that can do more with less data, getting benefit from 'leaner' and more focused data and not further encouraging a situation where people feel that they can create and store as much data as possible. The goal should be focused on how technology can enable us to create the balance towards innovating and using data responsibly.

Overall, we need to think about responsible data use in the fullest sense, taking responsibility not only for how it is used, but also how it is maintained, kept accurate and representative, making sure it is used appropriately for the public benefit and ensuring that public benefit is robustly understood and followed while involving citizens.

What's missing or needs development to deliver responsible data?

Define clearer outcomes

The National Data Strategy would benefit from a clearer articulation of the vision and the strategic direction it is trying to achieve for responsible uses of data. There are questions about bringing more clarity around outcomes, challenges, trade-offs and potential harms. The strategy would benefit from acknowledging the overarching systems and cultures around the use of data, and shifting the focus from data itself, which narrows down the perspective. This would have the potential for the government not only to become a model data processor, but also to be a leader in developing novel data-governance architectures.

Data is not responsible, people are

The National Data Strategy talks about data for a 'fairer' society – a concept that does not have a clear definition – and leaves open questions around who has responsibility to ensure data is used ethically and responsibly. Potentially, a chief data officer could be considered for ensuring responsible uses of data, as long as the appropriate levels of accountability for this position are ensured. There is currently no mention of power dynamics between companies, users and governments, or of the ethical stance actors need to take themselves. For example, how do we address the power of the private companies that penetrate the public sector? This leaves a gap in the strategy, one that can potentially be addressed by robust governance, clear redress mechanisms, and firm and timely interventions from regulators.

Public engagement

The National Data Strategy commits to undertaking a national engagement campaign on the ‘societal benefits of the use of government data’ and articulates the need to ‘increase public engagement’. However it is not clear in the National Data Strategy how engagement will empower citizens to influence decision-making or strategies related to data.

Citizens are the subjects of algorithmic decision-making, but the National Data Strategy does not indicate how to integrate algorithmic accountability into everyday life: how citizens will be part of processes for data and algorithmic scrutiny, what people’s routes to understanding data-driven decision-making looks like, or what recourse is available. Often minority communities are the target of algorithmic decisions, and there is currently no clarity or emphasis on how minority communities are going to be involved in decisions that disproportionately affect them.

Value stance

The National Data Strategy contains language around ‘value-based’ and ‘ethical’ use of data, but little elaboration on what that details: which values, why those values and inherent trade-offs between values.

Explanation about how algorithmic systems work is important, and information needs to be tailored to the audience. However, efforts to move towards deciphering AI fail to address concerns around what values are being propagated through these systems. There is a need to articulate clearly what values society as a collective believes are important, and then hold companies and government accountable to building systems to those standards.

For example, legislation could set precise public policy goals that companies need to meet in order to protect against discrimination, health misinformation or election manipulation. All necessary and proportionate measures to achieve these goals would need to be considered, and an independent regulator with a clear mandate to demand access to all relevant data and the ability to perform supervisory tests and monitor due diligence obligations would be needed.

Addressing the question of values also means a shift in mentality from ‘Can I build it?’ to ‘Should I build it?’, which needs to be encouraged by policy and at the same time economically incentivised by market practice. Another recommended change in perspective is represented by understanding automated decision-making processes as sociotechnical. Such decisions should not fall on a single technical department, but should be viewed more interdisciplinarily – and potentially codes of conducts could be developed for data scientists.

The strategy emphasises the need for innovation, but safeguards and redress mechanisms are not mentioned. The recommendation is to include demonstrable and enforceable accountability programmes for organisations (public and private) to implement data-management policies, procedures and controls leading up to a safety-first implementation. For example, there could be an end-to-end view on algorithmic accountability, from data to logic to authorisation.

Further topics discussed

Synthetic data

The National Data strategy says that ‘there may also be value in creating and sharing synthetic data to support research and innovation, as well as other privacy-enhancing technologies and techniques’. Concerns were expressed that this could be interpreted as a way for big synthetic datasets to be developed and ‘rented’ for widespread world use and universal research application. This type of top-down approach would not be the most appropriate way forward. Instead, a hypothesis approach should be considered, which argues for using the smallest synthetic dataset needed in order to achieve intended project goals and test initial assumptions.

International data transfers

The National Data Strategy begins its section on international data availability with the statement: ‘It is hard to overstate the importance of flows of data across borders to support economic development and global cooperation.’ However, some challenged this assumption, arguing that free data flows can lead to the concentration of data in the hands of a few dominant players, making it more cumbersome for SMEs to access high-quality data. Another concern was that the focus on international data flows may distract from the problem of setting up well-functioning data ordering regimes. New and innovative mechanisms would advocate for value-preserving data governance schemes, which translate into data being available for generating benefit for individuals and society at large.

The National Data Strategy commits to ‘supporting international data flows while ensuring that transfers of personal data from the UK uphold high data-protection standards’. Our workshop conversations highlighted that without serious normative scaffolding, especially in terms of human and consumer rights, the international flow of data in a global data ecosystem can also cause harm. For example, the digital services that come from outside the UK may not conform with domestic consumer protection, they may use unfair business practices and undercut data privacy standards or safeguards against algorithmic decision-making.

Furthermore, on the future of adequacy regime after Brexit, injecting free data flows in trade law is a questionable way forward, because this tends to crowd out other efforts around multi stakeholder governance that aspire towards a more rights-preserving and inclusive approach to safeguarding human rights. Instead, a more viable strategy would be, as mentioned above, to create a digital rights scaffolding, together with liberalising cross-border data flows while respecting data protection. Important judgements from the Court of Justice of the European Union ([Privacy International](#), [La Quadrature du Net](#) and [Schrems II](#) rulings) are creating greater difficulties in taking on EU adequacy agreements, and academics and legal professionals are still in the process of assessing which personal data flows are directly affected by the bulk interception regime in the UK.

The key message was that instead of picking the most attractive aspects of international data flows, there is also the need to contribute to a high level of protection of personal data and individual rights.

Digital IDs

Digital IDs were discussed as offering advantages for citizens, but in doing so requiring massive data linkage of discrete public-sector datasets, which can be used for good purposes such as research and statistics, but also for profiling behaviour. The result is a 'transparent citizen'. Linking data is increasingly used for predictions and risk indicators in social services, with algorithmic decision-making being put into question.

It was argued that any move toward a digital ID should explore capabilities for developing virtual identities where citizens have different identities in different contexts, and only the citizens can link them. This option can be implemented on the individual's device, therefore limiting privacy risks and leaving open the possibility of contributing data for research.

Resources and examples mentioned during the event:

- [Data Protection Impact Assessments as rule of law governance mechanisms](#)
- [ICO Innovation Hub project report](#)
- [Guidance on the AI auditing framework](#)
- [Project ExplAIIn](#)
- [City perspectives on digital rights](#)
- [Proposal for an 'all-in-one' algorithmic impact assessment](#)
- [Better machine learning through data minimisation](#)
- [Data and democracy](#) [event recordings]
- [When race/ethnicity data are lacking: using advanced indirect estimation methods to measure disparities](#)
- [A toolkit for centering racial equity within data integration](#)
- The Alan Turing Institute is working with Camden City Council citizens to develop a Data Charter to consider how data should be used and implemented.

About us

About the Ada Lovelace Institute



The Ada Lovelace Institute was established by the Nuffield Foundation in early 2018, in collaboration with the Alan Turing Institute, the Royal Society, the British Academy, the Royal Statistical Society, the Wellcome Trust, Luminare, techUK and the Nuffield Council on Bioethics.

The mission of the Ada Lovelace Institute is to ensure that data and AI work for people and society. We believe that a world where data and AI work for people and society is a world in which the opportunities, benefits and privileges generated by data and AI are justly and equitably distributed and experienced.

We recognise the power asymmetries that exist in ethical and legal debates around the development of data-driven technologies, and will represent people in those conversations. We focus not on the types of technologies we want to build, but on the types of societies we want to build.

Through research, policy and practice, we aim to ensure that the transformative power of data and AI is used and harnessed in ways that maximise social wellbeing and put technology at the service of humanity.

We are funded by the Nuffield Foundation, an independent charitable trust with a mission to advance social well-being. The Foundation funds research that informs social policy, primarily in education, welfare and justice. It also provides opportunities for young people to develop skills and confidence in STEM and research. In addition to the Ada Lovelace Institute, the Foundation is also the founder and co-funder of the Nuffield Council on Bioethics and the Nuffield Family Justice Observatory.

Find out more:

Website: [Adalovelaceinstitute.org](https://adalovelaceinstitute.org)

Twitter: [@AdaLovelaceInst](https://twitter.com/AdaLovelaceInst)

Email: hello@adalovelaceinstitute.org

About the Centre for Public Data

THE
CENTRE
FOR
PUBLIC
DATA

The Centre for Public Data is a new, non-partisan organisation with a practical mission to improve the UK's public data. We work with partners to include data-focussed provisions in new legislation, and campaign to fill important gaps in data collection.

Find out more:

Website: centreforpublicdata.org

Twitter: [@CFPData](https://twitter.com/CFPData)

Email: contact@centreforpublicdata.org

About the Institute for Government

INSTITUTE
FOR
GOVERNMENT

The Institute for Government is the leading think tank working to make government more effective.

We provide rigorous research and analysis, topical commentary and public events to explore the key challenges facing government.

We offer a space for discussion and fresh thinking, to help senior politicians and civil servants think differently and bring about change.

We believe the quality of government data, and the better publication and use of government data, is vital for government effectiveness (government needs to understand how it is working and how this can be improved) and for accountability, allowing parliament, the press and the public to scrutinise it.

Find out more:

Website: instituteforgovernment.org.uk

Twitter: [@instituteforgov](https://twitter.com/instituteforgov)

Email: enquiries@instituteforgovernment.org.uk

About the Open Data Institute



The Open Data Institute works with companies and governments to build an open, trustworthy data ecosystem, where people can make better decisions using data and manage any harmful impacts. theodi.org

About the Royal Statistical Society



The Royal Statistical Society (RSS), founded in 1834, is one of the world's most distinguished and renowned statistical societies. It is a learned society for statistics, a professional body for statisticians and a charity which promotes statistics, data and evidence for the public good. Today the RSS has around 10,000 members around the world. rss.org.uk

Getting data right: perspectives on the UK National Data Strategy 2020 is published by the Open Data Institute

© 2020 Ada Lovelace Institute, Centre for Public Data, Institute for Government, Royal Statistical Society, Open Data Institute.