# National Data Strategy Consultation Response

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

Deadline for responses 2 December 2020. Submitted 30 November 2020.

## Summary

This response has been written by Tom Forth (Head of Data ODI Leeds) with input from the sponsors of ODI Leeds and others. This document was open on the web for over two months and received comments and suggestions from over a dozen people.

ODILeeds is a not-for-profit organization with about ten employees and frequently-employed associates. It is funded in part by its sponsors listed at https://odileeds.org/services/sponsors/ with the majority of income from events and paid work.

The response urges the UK government to,

- More thoroughly consider the costs of central government imposition of standards and regulations, especially in relation to local and devolved sovereignty and budgets. Many of the case studies of success in the report were successes because they exercised local sovereignty and opted out of the national control and regulation that this report now proposes more of. Centralisation may be the correct decision, but the drawbacks need considering.
- Be more realistic about the UK's position within the world since Brexit. Brexit is unlikely to be "done" in our lifetimes. The EU is the world's only major regulatory power on data. There is room for two more such powers: the USA and China. The UK will not be such a power and will need to follow more than it leads in most areas. This is an opportunity for the UK government to do less.
- Focus on the biggest wins first by urgently creating an open postal address system for the whole UK.

## Overview

Why should local government data on waste collection interact with a franchised train company's timetable? How could any strategy relevant to both hope to provide any useful detail or guidance to either?

Thinking through those two questions is why it is not obvious that the UK should have a national data strategy.

This UK National Data Strategy is an important first argument for one. Its tone is a welcome balance of positivity, optimism, and caution. It recognises the importance of data across business, society, and government to increase productivity, improve lives, and strengthen our country. It forms a strong basis for the wider discussion that it invites throughout.

The case studies are strong. Transport for London's excellence in the collection, use, and safe release of data is world leading. It is the result of decades of hard, controversial, and innovative work. Much of this innovation was enabled by local sovereignty and associated opt outs from national standards and laws.

The NHS' health records are unrivalled internationally in their coverage of a large and diverse population. They are the foundation for significant scientific and industrial strengths that other countries will be unable to emulate for decades, if ever. It has also enjoyed significant autonomy from direct central government control. It enjoys greater public trust as a result.

The ONS and the ONS Data Science Campus have proven themselves excellent at innovating and interacting with people, governments, businesses, and institutions right across the UK. Again, much of this achievement is founded in their distance from government and their exemption from central government standards and control. For example, the ONS still publishes data in Excel spreadsheets having identified that as a user need where the UK government forbids publication of data in Excel spreadsheets, having identified that as a user need.

The first major concern in light of these three examples is that the strategy does not consider sovereignty. Specifically it seems to advocate the central ownership of (and hints at the central imposition of) data standards, ways of working, and interoperability requirements. The potential upsides as measured by efficiencies of scale and reduced duplication of effort are obvious and large. The costs in terms of lost resilience, less opportunity for innovation, costs of complying with rules and implementing standards, and greater distance between government and the people do not seem to be considered.

These costs were obvious before 2020. Covid-19 has made them unmissable.

Beyond the Covid-19 response, which we refer to in response to questions later, this trade-off is widely discussed in the devolved nations and in the regions of England. We have already mentioned the success of TfL with data, a success which was enabled by local sovereignty and exemption from national standards and laws. The clearest manifestation of this opt out is that London's Oyster card does not use the ITSO national standard for public transport smartcards as used elsewhere. This decision and its ability to move more quickly contributed to its success and should be considered deeply. We can estimate the potential cost of standards by comparing the progress London made on smart ticketing by opting out of them compared to the progress other places made within them.

Other examples of the benefit of retaining local control include the decision of the devolved nations to collate their own census returns and create their own indices of multiple deprivation. This leads to significant advantages. Specifically, Scottish IMDs are reported at datazone resolution and Northern Irish IMDs are reported at SOA resolution, both more detailed geographies than English and Welsh IMDs. And because each nation's IMDs report the rank of a place within each nation the smaller nations' IMDs are more useful for reporting local variation of deprivation than IMDs in England. Analysis of deprivation within English cities and regions is far too often poor because an English national IMD is used instead of a local one. The gains from imposing a national and standardised approach in making it possible to do whole-UK analysis of deprivation more easily should be weighed up against the cost that the data will be worse.

The second major concern is that the strategy underplays the challenges facing the UK as a result of the UK leaving the EU. The UK government signed a withdrawal agreement in 2019 and won an election on a promise to implement it. In September 2020 the UK government admitted that it is willing to break the law to overrule parts of that agreement which it now believes to be flawed.Those flaws were pointed out at the time and ignored. There are similar flaws in this strategy and the government should not ignore them again. The UK government will work most effectively with partners throughout our country and the

world if it is seen to understand the challenges we face. What is widely known as mere bravado in Westminster and Whitehall is rarely understood as so outside of it. A more realistic evaluation of the UK's power on data regulation would be useful for gaining international credibility and as a foundation for internal collaboration.

The third major concern is that the strategy does not seem to identify a critical weakness in the UK's national data infrastructure. Specifically that the UK lacks an open postal address system, leaving both business and government services behind other countries.

# Response to questions

The strategy asks 19 questions which will be usefully answered.

Q1. To what extent do you agree with the following statement: *Taken as a whole, the missions and pillars of the National Data Strategy focus on the right priorities.* Please explain your answer here, including any areas you think the government should explore in further depth.

The four pillars and five missions are broadly correct. Pillar 1 and 3 could be improved by focusing more on the simple point that **data must exist and its existence must be proven for it to have value.** Even where it is responsible for data collection and publishing, central government continues to prefer talking about what data might be useful and in what format it should be collected, stored, and shared ahead of actually collecting the data and sharing it.

Both the pillars and missions fail to mention sovereignty and there would be value in considering this more widely. The ownership of data, the right to collect data, the choice of standards and formats for storing and sharing data, and the right to withdraw consent to use data are questions of sovereignty. The sovereignty is of citizens, of companies, of organisations like the NHS, of different departments within government, of different tiers of government within the UK (see the TfL example already given), and of different countries both within the UK and around the world.

Sovereignty gives a right to experiment without permission and is thus a fundamental source of innovation. Sovereignty is often what is sacrificed for the more obvious good of interoperability, standards, and working together. This fundamental tradeoff does not seem to be explored in the strategy.

Q2. We are interested in examples of how data was or should have been used to deliver public benefits during the coronavirus (COVID-19) crisis, beyond its use directly in health and social care. Please give any examples that you can, including what, if anything, central government could do to build or develop them further.

- The French government's simpler, more complete, and higher quality open data on Covid-19 infections and hospitalisations in each of the 101 départements of France and its display on a publicly available dashboard linking back to the source data has increased public trust, reduced public fear, and accelerated economic recovery. That the same data remains overly complicated, incomplete, and of poor quality for England is important for the government to accept. If the UK government does not understand the depths of the failure of its current digital and data strategy in relation to Covid-19 it has little hope of achieving the goals in this strategy. It will be particularly important for the UK government, despite far greater resources, to accept that its data response was poorer and remains poorer than that in Wales and Scotland and to understand why.
- French government data on furloughing and government assistance for every département has greatly improved national debate on the correct economic measures to take in recovery. It has

helped businesses stay confident to continue trading and helped local and regional governments and government institutions to target support. The lack of equivalent data in the UK is a reminder that great government data cannot just come via the ONS, that administrative data directly published by government departments within the UK can be extremely valuable, and that we are currently quite poor at publishing it.

- Private sector data such as Indeed job listing and Google Mobility has consistently been better, more timely, and more widely available than similar data which exists within the UK government. It would be worth exploring the reasons why companies do not share such data more often and work on reducing any barriers, particularly legal ones, to such sharing.

Q3. If applicable, please provide any comments about the potential impact the proposals outlined in this consultation may have on individuals with a protected characteristic under the Equality Act 2010?

Q4. We welcome any comments about the potential impact the proposals outlined in this consultation may have across the UK, and any steps the government should take to ensure that they take account of regional inequalities and support the whole of the UK.

The UK national data strategy is a centralising prospectus. In its own words it is **"a whole-government approach driven from the centre"**.

This is in keeping with the last ten years of government spending within England (see ONS public sector employment headcount and the decline of local government) and this government's approach to digital government and data more widely (for example the centralisation of voter registration, government notifications, and UK government website domains by GDS).

Centralisation has many positives and it is right that the proposals celebrate them. But it also almost inevitably leads to an increase in regional inequalities when the centralised approach and so much of its associated spending and decision-making happens in the strongest regional economy of the UK.

It is mostly not through direct spending that such a concentration of power and spending creates regional economic inequality. Rather it is the well-observed property for proximity to generate agglomeration benefits. Specifically, if the UK's data policy is debated and set in London then the requirements and concerns of businesses in and close to London will be weighted more heavily in that policy. Businesses with their senior team elsewhere will be disadvantaged.

Even where processes for setting data standards and setting data policy are decentralised and open, there are considerable costs to engaging. Local government and small businesses are at a great disadvantage to central government and big businesses in any open standards setting system due to their lower budgets and the fact that ultimately the centre will choose. Where poor systems are designed centrally, such as data.gov.uk, the right to opt out locally is important to avoiding costs being imposed locally.

Short of significant decentralisation of power and a sustainable local government funding settlement (which seems beyond the scope of this report) two approaches are likely to work in reducing the negative side-effects of centralisation,

1. The location of major parts of the UK government responsible for data policy outside of London. This was a request of multiple respondents to the Centre for Data Ethics and Innovation consultation and has since been ignored. The creation of further authorities, centres, and

institutes in London does not inspire confidence that this will ever change, but this is a win that is ready to grasp.

2. Central government doing less. Quite bluntly, the less central government does, the less it does in London, and the less the comparative advantage to businesses in London will be. The fewer standards that are imposed centrally, the less chance that they will not meet needs locally or impose costs locally.

A third approach has some potential but comes with a cost that probably makes it unwise.

Specifically, the UK government has in recent years tried to stimulate cash-strapped local government innovation in digital and data with ring-fenced funds or innovation prizes. This money often comes with conditions attached that impose central government standards and ways of working (user-centred design, discovery phases, use of centrally-defined open standards, etc…) and are available only for projects that meet the central government's identified needs, not the needs of the local government. This erosion of local sovereignty is often explicitly weighted in evaluations of bids as "cultural fit" and typically wrapped up in the language of "user needs" with the argument that a central approach that explicitly focuses on these needs will be superior to a local approach that does not, or does so differently.

By reducing local sovereignty in this way, much of the potential to innovate with data is lost and where money is spent by local governments it is often spent on problems that are not locally relevant. Without a proper funding settlement for local government these small pots of funding struggle to generate sustainable innovation and support business growth.

Q5. Which sectors have the most to gain from better data availability?

We don't think anyone knows and we'd be highly sceptical of anyone who thought they did.

Q6. What role do you think central government should have in enabling better availability of data across the wider economy?

The biggest thing that only central government can do and which would have a huge impact is to fix addressing in the UK. The UK currently lacks an open address system. Specifically a list of all UK premises, both residential and postal, with the street address and postcode is not available under an open licence. This imposes very significant costs on small businesses and reduces the quality of address data to every business. It significantly reduces innovation in digital services within the UK and puts British businesses at a disadvantage compared to European rivals. It greatly limits data that local government can publish openly and software that local government and its suppliers can publish openly.

In other areas, central government has a crucial role. Legislation such as GDPR gives individuals rights over the data held about them and is best done on at least a national scale. Open banking makes secure access to data in standard formats via an API a right for customers, stimulating innovation in the financial services sector. Again, this is best done on at least a national scale. Government released open data generates innovation and efficiencies across society and the economy. Where data is collected by central government this is the right scale to release it. The UK government could be much better at releasing administrative data than it is today.

In all these considerations central government should not forget that all regulation imposes direct costs (compliance) and indirect costs (sovereignty and the right to innovate) on those it is regulating. It is clear from their omission from the strategy currently that the indirect costs are not being considered. If they are, it is more likely that central government will find the right balance between action and inaction. One

specific example to consider is the extent to which Google were within their legal rights to collect, analyse, and then distribute its Google Mobility rankings during the Covid-19 pandemic. If they were not (users had certainly not opted in to sharing this data for its ultimate purpose of tracking pandemic response) then the government should work to ensure that such value is not forbidden in future.

It should be the ambition of central government that all data it collects should be shared openly, as soon as possible. If this is not possible data should be shared back to the people collecting it in the first place so that they can use it effectively. There are numerous examples within the NHS and Public Health England where hospitals provide detailed data to central government bodies but cannot themselves access data once it has been collated across the whole system.

Q7. To what extent do you agree with the following statement: *The government has a role in supporting data foundations in the wider economy.* Please explain your answer. If applicable, please indicate what you think the government's enhanced role should be.

Government -- local, central, and devolved -- have a key role in supporting data foundations in the wider economy. This includes by,

- Regulating the use of data (things like GDPR).
- Releasing data for wider use (usually as open data since we know this generates most use and value) as part of the functioning of government and when it awards contracts and franchises to others.
- Directly funding innovation with data via quasi-independent research institutions, universities, etc...
- Funding training for people and businesses to work with data.

They must balance this positive with the costs of imposing regulation on other tiers of government and on businesses.

The UK government has an opportunity to lead by example in supporting the data foundations of the wider economy by promoting openness in its work on data. Specifically, it could redouble its efforts to be radically open in what it does by using the web as it was designed for the communication and delivery of its data strategy. Government organisations should have websites from the day they are created, with regular updates on their work, the data that they are using to take and justify decisions, and links to other organisations that they work with.

Q8. What could central government do beyond existing schemes to tackle the particular barriers that small and medium-sized enterprises (SMEs) face in using data effectively?

Central government should be wary of any intervention in this area beyond incentivising investment by companies more broadly. UK companies of all sizes have long invested less than companies in Europe and North America and this trend has strengthened in the last five years. Using data effectively in a business requires investment and there is little evidence that investment in using data is not part of the wider trend of weak business investment in the UK.

One area where government could intervene to increase business investment is through local mentoring schemes and short-duration training schemes. This type of action will be better delivered via local government than by central and is therefore unlikely to be relevant in a national strategy.

Q9. Beyond existing Smart Data plans, what, if any, further work do you think should be done to ensure that consumers' data is put to work for them?

Q10. How can the UK's data protection framework remain fit for purpose in an increasingly digital and data driven age?

The current system works pretty well and does not require major change. The UK will need to respond to adequacy requirements in trade deals we sign. We must be ready to respond, but changing in advance seems wasteful.

Q11. To what extent do you agree with the following statement: *the functions for the Centre for Data Ethics and Innovation (CDEI) should be Artificial Intelligence (AI) monitoring, partnership working and piloting and testing potential interventions in the tech landscape?*

The CDEI seems to have largely ignored the opinions it received in response to its founding consultation. Except for one meeting in Edinburgh it has never held a meeting outside of London. It seems to have only just created a website, which remains hard to find and does not project any independence from the UK government. There is little knowledge of its existence. Even to those of us who are interested in it, it remains unclear why it was set up and what it does.

This and the other institutions mentioned in the strategy i.e. Alan Turing Institute, the National Innovation Centre for Data, the ODI, the Data Skills Taskforce, the AI Council, the UK Cyber Security Council and others should be measured on how they lead by example in terms of data publishing, sharing and use of data. The government should regularly check that their location represents, and can work with, a data industry that is widely distributed across the UK.

Q12. We have identified five broad areas of work as part of our mission for enabling better use of data across government:

- Quality, availability and access
- Standards and assurance
- Capability, leadership and culture
- Accountability and productivity
- Ethics and public trust

We want to hear your views on which of these actions will have the biggest impact for transforming government's use of data.

Availability and access are the most important. The government should consider that efforts to improve quality will likely reduce availability and access at least in the short term. This should be avoided where possible, even if it means that achieving quality takes longer.

Bad data in a bad format is better than no data. The government's Covid-19 response is a useful example of how aiming for excellent quality data in the best format often results in no data for months and then a poor result nevertheless. The UK's Covid-19 dashboard, based on open standards, with a complex API, is in many ways technically excellent. But it was months late, remains unreliable, is still missing key simple data in favour of largely worthless complex data, and is difficult for most people to work with.

A Tableau dashboard with a spreadsheet download button would probably have been cheaper, better and would have been available much earlier. It would have also served as an example to local

government and other institutions that cannot afford to develop custom dashboards for their needs, but could learn from central government using widely available tools.

In thinking about user needs with data and digital it is unclear that the UK government is currently thinking about its population and partners today ahead of the purity of its technology strategy.

Q13. The Data Standards Authority is working with a range of public sector and external organisations to coordinate or create data standards and standard practices. We welcome your views on which if any should be prioritised.

In all of our discussions in preparing this response, none of dozens of people had heard of The Data Standards Authority. Until recently it had no website. Now it has a website as part of a government department that tells us almost nothing about it. We do not know where it is based, who works for it, who sits on its steering board, who makes up its peer review group, or who leads it. We have no idea how we would engage with its work or what data standards it might impose on us.

It would increase trust in government in this space if it and other similar institutions and units of government worked either in a more modern and open way (get a website from day one and tell us what you're doing on it) or a decentralised way (so that we might know somebody involved).

Q14. What responsibilities and requirements should be placed on virtualised or physical data infrastructure service providers to provide data security, continuity and resilience of service supply?

Q14a. How do clients assess the robustness of security protocols when choosing data infrastructure services? How do they ensure that providers are keeping up with those protocols during their contract?

Q15. Demand for external data storage and processing services is growing. In order to maintain high standards of security and resilience for the infrastructure on which data use relies, what should be the respective roles of government, data service providers, their supply chain and their clients of such services?

Q16. What are the most important risk factors in managing the security and resilience of the infrastructure on which data relies? For example, the physical security of sites, the geographic location where data is stored, the diversity and actors in the market and supply chains, or other factors.

Questions 14 to 16 are all about security of data infrastructure. This is a combined answer.

For data infrastructure crucial to national security and personal safety a more significant and expert response is required than we can give.

Outside of that scope, the requirements of GDPR are pretty good and in most cases any extra responsibilities and requirements would be counterproductive.

Specifically, existing schemes such as Cyber Essentials Security should be studied and their flaws avoided. The scheme itself forbids public discussion of its flaws, so they are not widely known. There may be many further such schemes that we are not aware of as a result of similar censorship. The process is considered counterproductive (it reduces their security instead of improving it) and costly by many companies who go through it. Security through secrecy is not a good strategy.

The requirement by parts of the UK central government and its agencies for contractors to gain certification is an excellent example of how well-intentioned regulations by central government can

impose large costs on UK businesses, price out smaller companies from government contracts, and reduce innovation with data more widely.

There are early signs that similar thinking in the areas of algorithmic accountability are imposing very significant costs too. The government must be aware that a large industry exists to lobby for greater regulation and debate on these topics, which those doing the lobbying often then provide. Such understanding is widespread within the competition and markets authority and advice should be taken from the experts there on how to avoid excessive regulation and regulatory capture as well as insufficient regulation.

Q17. To what extent do you agree with the following statement: *The government should play a greater role in ensuring that data use does not negatively contribute to carbon usage*?

Somewhat agree. The UK national government has a role to play in all areas of reducing carbon emissions. In relation to data centres and cloud computing there is a global gap for trustworthy data on the carbon cost of these industries. Third parties have proven themselves untrustworthy in both directions. A UK national institution supported by central government could reasonably expect to be the world leader in quantifying the carbon cost of data centres and cloud computing and lead the conversation on reducing and taxing those emissions.

Q18. How can the UK improve on current international transfer mechanisms, while ensuring that the personal data of UK citizens is appropriately safeguarded?

We have not heard any arguments that suggest that the UK is in a good position to lead on this. If we are not, we would do better to spend our money elsewhere and maintain limited competences in this area.

We will seek EU ‘data adequacy’ to maintain free flow of personal data from the EEA and we will pursue UK ‘data adequacy’ with global partners to promote the free flow of data to and from the UK and ensure it will be properly protected.

Q19. What are your views on future UK data adequacy arrangements (e.g. which countries are priorities) and how can the UK work with stakeholders to ensure the best possible outcome for the UK?

We are open to this arrangement if those promoting “UK data adequacy” can clearly give examples where this would help UK businesses more than the cost of dealing with double regulations. Until then, the UK should aim for EU data adequacy above all, even if our efforts are not initially recognised. Unilaterally emulating the EU’s regulations on data would be a cheap and easy option that would reduce the need for extra bureaucracy within the UK.