medConfidential’s detailed read of the DCMS’ National Data Strategy...

The environment for implementing the strategy, and next steps

'Data across government' should stay with GDS in Cabinet Office, which can influence behaviour across the Departments – but if what we want as a country is just, fair and trustworthy use of data, then the **ICO should return to the Ministry of Justice as the departmental sponsor**. DCMS wishes to promote data, but the strategy fundamentally conflicts with the quasi-judicial role of the ICO.

The CDEI should be restructured according to the norms being set for the Senior Data Governance Panel in MoJ,[1] including most notably lay members (as the DHSC National Information Board had). Such a broad source of advice and input would increase the assurance across Government that the data views any departmental project is getting have a solid grounding beyond what a project team is willing to hear.

Next steps in implementing

Identifying accountable owners for each action is a good first step; on what basis, frequency and to whom will they report?

> *It is also likely that there will be **wider public interest in the social aspects** of the strategy.*[2]

This is very likely and, though they are mentioned in some parts, there is too little emphasis placed on social dynamics throughout; data does not exist in isolation, and its use by different actors can have radically different outcomes. What's being talked about are some profound – and some rather less profound – *technosocial* interventions.

It would therefore help if DCMS would provide some indication of its *vision*, i.e. what might the UK look like if what it espouses in all of these "Pillars" and "Missions" is successful? It's all very well to talk in terms of "Opportunities", and eminently sensible to engage in ongoing "Monitoring and evaluation" – but **what does good look like**, where is it we are actually heading, and when will you publish your metrics?

This consultation is likely to elicit responses from the usual suspects and vested interests, so how is DCMS proactively seeking the views of those most likely to suffer adverse consequences?

While customers may be able to choose a different commercial supplier if they are dissatisfied with the service they receive from one provider, the same is not the case for citizens when dealing with government or the public services. The power relationships between corporations and customers and citizens and the database state are very different; expectations are different too. And it is the most marginalised and vulnerable who often pay the heaviest price for policies and services designed to fit Governmental notions of the 'norm'.

[1] Our early thinking on this model started with major reforms deriving from the various Justice Councils inside MoJ. However, the recent announcement by HMCTS/MoJ of the Senior Data Governance Panel - https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/925341/HMCTS_Making_the_most_of_HMCTS_data_v2.pdf - is a much stronger and more structured base for informed debate and competent advice on information governance issues, subject to full implementation over the next few months. Any problems in implementation can be reformed over time given the potential strength of the SDGP structure, if balanced appointments are made.

[2] Section 8, 'Who we are seeking to consult with?'

Blow-by-blow

Given the lack of both pagination and paragraph numbers in the strategy document as published, the comments below are collated under the sub-headings given within the document.

Ministerial foreword

Sharing information "quickly, efficiently and ethically" is not the same as doing so *lawfully*. Nor is it a guarantee that such sharing will result in anything *effective*. Taking powers and making promises (or claims) is one thing; delivering on them is something completely different. If what you are doing involves data, however – assuming you are doing it properly – what should always be possible is to provide evidence.

If your "high watermark" is the untransparent mass use of patients' data, stepping outside the statutory obligations of the Act under which the COPI Regulations and Notices were issued, ignoring and denying patient opt-outs rather than respecting the public health exemptions already in place, then you may end up leaving a ring of scum...

Who says "data use" is a "threat"? What we oppose is data *misuse* and *abuse* – are these <u>not</u> "threats against which to be guarded"? Is the government opposed to the consensual, safe and transparent use of citizens' data?

What does the government consider to be "its own data"? This is a more crucial question than it may at first appear, because a lot of the data the government holds is *citizens'* personal data – collected in the course of delivering or administering public services and functions, and *for* those purposes. Government cannot simply redefine such data as "its own", to do with as it wishes. That's not a "barrier"; that would be breaking the law.

What is the difference between being a "global champion of data use" and a 'champion of global data use' – the latter being implied by "encouraging the international flow of data across borders"? If this is truly the Government's goal, then why risk the UK's GDPR adequacy, which would definitely create barriers to the flow of data across borders?

"Underpinned by trust" is a fine goal – even a necessary prerequisite for success – so how does the strategy propose to achieve it? Trust does not just magically appear...

Executive Summary

“Data creates jobs” – except where it replaces them! How does the strategy propose to deal with automation? (The word appears just once in the document, in a footnote, despite Bank of England projections of tens of millions of jobs being automated out of existence – and that was before COVID...) Future economies are certainly going to be information economies, but you need to plan more for people than just a few hundred or few thousand data analyst jobs.

Painting a solely positive picture – “the use of data benefits us every day” – rather misses the point. What about the harms? The information ‘pollutants’ (and polluters) against which we must be vigilant if our information economy is to be resilient and survive, much less thrive? Relentlessly ‘selling the benefits’ without addressing the very real downsides is just poor strategy, as this and previous Governments have found time and time again.

It seems unlikely the UK will “influence” any global approach to data sharing and use as a single nation of 60 million; it cannot noticeably influence any single one of the Tech giants. (And while we’re at it, exactly how much VC has the UK attracted compared with, say, Silicon Valley in recent years?)

Is data really “non-depletable”? It has a ‘half-life’ and, as a resource, its value resides largely in being related to other things – the complexity and management of which come at a cost. While the same data may be used many times, not all uses are non-rivalrous [example...]

cf. Data misuse as “Missed Use”[3] blog post for “when organisations do not make good use of the data they already have”. Maybe rather than thinking of data solely as a resource, the Strategy should consider its impact as toxic waste as well?

[3] https://medconfidential.org/2020/data-misuse-as-missed-use/

Pillars (of effective data use)

Are (absence of) these things *really* what “prevent the best use of data in the UK”? How does DCMS define “best use” anyway? And what does it mean by “effective”, and effective *for who* – for Government? For citizens? For commercial interests?

- **Data foundations:** “fit for *what* purpose”? GDPR requires that processing of personal data is *purpose-specific*, so does this mean “anything the government or ‘innovators’ want to do” should be added as a purpose to every (public) collection of data? Something can only be ‘fit’ for a purpose that is clearly *defined* – up front – and any one purpose may have significant consequences for other purposes, e.g. where there is, and needs to be, an expectation of confidentiality. A strategy that focuses too closely on data itself and fails to attend to the people whom it is largely about, the context(s) in which it is generated and used, and the many competing interests and trade-offs is just building castles in the air.

- **Data skills:** so how about creating some genuine data-skilled pathways into and through different professions and careers? Don’t forget entry level – stop outsourcing to China and mTurk! – and also remember retraining is not only for those whose jobs have been or will be automated away.

- **Data availability:** “accessible, mobile and re-usable”? Pick two (i.e. 1 and 3)! If it wants a thriving data economy, the UK will invest in **safe settings** and get people to come – and keep coming – to us to use our citizens’ and country’s most valuable data. What BEIS and OLS apparently fail to appreciate is that once our data’s left the country, it’s gone. Are *they* going to get it back? The government and public services aren’t paying nearly enough for lawyers sufficient to negotiate the sort of contracts they hope to cut – if they can even be bothered to do so, cf. Hancock and Alexa. And, as we keep having to say, “In negotiations for patients’ data, the NHS sends doctors[4], while Google sends lawyers and trained negotiators”.

- **Responsible data:** it is wise of DCMS to acknowledge that innovation and research are not inherently “lawful, secure [medConfidential prefers the term ‘safe’, as in the 5 Safes], fair, ethical, sustainable and accountable”. And this is why it is fundamentally important that any strategy must attend to ensuring that all of these things are true – *and seen to be true* – of every use of (citizens’) data. Given the emphasis government puts on trust, maybe it should lead by operationalising **trustworthy** use of data? Anyone can *claim* to be “responsible”; those who are trustworthy *demonstrate* that they are.

[4] Or now maybe someone from the ‘Centre for Improving Data Collaboration’ (formerly the “Centre of Expertise”): https://www.nhsx.nhs.uk/key-tools-and-info/centre-improving-data-collaboration/ - a new business unit within the non-statutory NHSX, which seems to lack the commercial legal capacity to negotiate contracts over the use of millions, if not tens of millions, of patients’ medical records.

Actions (actually Missions)

There are dozens of ‘actions’ variously scattered throughout the document, over 60 of which are assigned to a Government owner in Annex A. Some of these are practical and specific, while others are far more vague and ‘aspirational’ – “commit to resolving long-running problems”, “work to better support” – and yet more focus on a particular action without any sense of managing impact or putting in place appropriate governance. Government can undertake a whole bunch of action(s), but without a clear, coherent and cohering vision, what will they achieve?

- **Unlocking the value of data across the economy:** a data strategy that only protects people’s “data rights” and businesses’ intellectual property is dangerously narrow; the focus cannot be solely on data, any strategy worthy of the name must have regard for, respect and engender human rights – and other rights too. We’ve seen this time and again in the failure of regulators to understand or even take into account laws beyond those for which they are responsible – such as common law confidentiality or the duty of care in health. The *first* duty must be to protect rights; only after those are secure can one consider how (or whether) to make data “usable, accessible and available across the economy”.

- **Securing a pro-growth and trusted data regime:** rather than constantly reaching straight for people’s personal data, the government should first *demonstrate* its trustworthiness (i.e. per Onora O’Neill, its competence, honesty and reliability) with non-personal data, e.g. operational data. There must be economic growth potential in this type of data, and solving knotty problems such as companies’ monopolistic ‘data-hoarding’ once they win public contracts would help prove that government is capable of properly managing ‘data sharing’ that involves personal data.
- **Transforming government’s use of data to drive efficiency and improve public services:** unfortunately, the benefits of automation and data use are not always evenly distributed between the citizen and the state. Government “efficiency” does not always result in a higher quality result for the citizen, and one must always ask who benefits from these “improvements” – officials and the machinery of state itself, or those they are supposed to be serving?
- **Ensuring the security and resilience of the infrastructure on which data relies:** so if the infrastructure is such a “vital national asset” why does the government, e.g. pursue a policy of “Cloud First”? Concentrating so much of the machinery of state in the hands of a few data-acquisitive US providers seems contradictory – not to mention the vast quantities of operational *metadata* that the government seems entirely content to hand over. A sensible data strategy would at least consider investing in the ‘sovereign data capacity’ to process our own key national assets.
- **Championing the international flow of data:** as the pandemic has shown, as have the decades before, international collaboration and research is entirely possible within current data protection regimes. As have been payment of people’s salaries, and connecting with loved ones. If the UK wishes to “champion” appropriate flows of data across borders and avoid “fragmented regulatory regimes” then maybe it would be a good idea not to deviate from GDPR, and to align UK’s data practices (several of which are already inappropriate) so as to achieve adequacy after Brexit?

1. About the NDS – including ‘What we mean by data’

This section acknowledges that this “framework strategy” focuses on *government’s* role in harnessing data – which is odd, given that role was taken back by the Cabinet Office shortly[5] before the publication of this much-delayed document.

Throwing one’s hands up and saying something is “hard to define” is not a sound basis on which to strategise! For Government policy, it would have been more appropriate – at a very minimum – to collate all of the relevant statutory definitions. These are things and terms that can at least be agreed, which can thereby locate the strategy in existing legal frameworks and identify precisely where changes may be required. If indeed they are.

Years more talking about “overcoming barriers to data sharing”, etc. without defining precisely what and where those “barriers” are – or claiming to want to “increase public trust” without specifying the safeguards and rules that will be enforced – is a hollow exercise, and will achieve neither.

Also, as regards definitions: once you have agreed what something *means*, there is still the job of being able to ascribe relative value to each thing. This does not necessarily mean a monetary value – that is proving problematic enough! – but some way of prioritising, e.g. public value, when (as every strategy must) it comes to trade-offs.

When thinking about “government's own data use”, and the nature of data that the government collects in the course of delivering services, the National Data Strategy must clearly attend to – and have regard for – both personal data and special category personal data. It is not good enough to ‘hide behind’ vague, often non-statutory terms like “administrative, operational and transactional” data, or the notion that the government somehow magically gets statistics without first having processed the underlying individual-level data – be that personal data, business microdata, or whatever.

When DCMS says “for many of the issues highlighted in this strategy a number of questions remain unanswered” what <u>are</u> those questions? What “further research and analysis” <u>is</u> required? Saying “We don’t know some stuff” without saying what it is you don’t know, and how you propose to find out, is less than confidence-inspiring. Building an evidence base is good, but the process must be open and transparent if it wants to gain (and retain) public trust.

[5] https://questions-statements.parliament.uk/written-statements/detail/2020-07-22/HCWS417

2. The data opportunity

To quote BEIS[6] in 2019's "Regulation for the Fourth Industrial Revolution":

> *The Fourth Industrial Revolution is of a scale, speed and complexity that is unprecedented. It is characterised by a fusion of technologies – such as artificial intelligence, gene editing and advanced robotics – that is blurring the lines between the physical, digital and biological worlds.*

Data clearly plays a fundamental role in this process; it can be a driver ('fuel' for commerce), a by-product ('exhaust', with possible polluting effects), a consequence (e.g. inferred data), and many other things. While section 2 does partly acknowledge this, it is vital to acknowledge that there is no "data opportunity" that can be abstracted away from all of the other parts. This strategy is about **technosocial** policy and practice, where the physical, digital and biological may indeed be 'fused' by and through their translation into data – but where the characteristics and constraints of each must be respected.

Just because, e.g. someone's genome is available as a digital data file, rather than as physical DNA inside their cells, does not mean that such derived genomic data can be divorced from the unique and uniquely sensitive real-world relationships and properties it represents. Its use in one context (e.g. consensual, ethical health research) may be publicly acceptable; the use of *the same data* for, say, criminal forensic purposes may make the collection and use of such data (even for those otherwise acceptable research purposes) far less publicly acceptable.

No coherent national strategy can afford to look at just 'data' in isolation; if the government wishes to engage with novel processing of novel types of data, it must be willing to draw some (statutory) red lines or risk losing public confidence altogether. It's not good enough just to say "We won't do that (for now)" or "We don't intend to do this" – public trust *requires* that the government put certain (re)uses legally off-limits if we as a nation are to be able to benefit from trustworthy, legitimate uses.

Where it is claimed there are "unnecessary barriers", what exactly are they? Or, put another way, maybe the government should define what it considers to be the *necessary* barriers?

The single word "opportunities" listed in the diagram in the executive summary are unpacked later in section 2, as follows (albeit with "Research" and "Public Services" in a different order):

Growth	Boosting productivity and trade
Jobs	Supporting new businesses and jobs
Public services	Driving better delivery of policy and public services
Research	Increasing the speed, efficiency and scope of scientific research
Society	Creating a fairer society for all

These all seem fine aspirations – but what, e.g. does "better delivery" mean? Better for who? For government, or for citizens?

[6] https://www.gov.uk/government/publications/regulation-for-the-fourth-industrial-revolution/regulation-for-the-fourth-industrial-revolution

While it is one thing for government systems to deliver a service (or policy) effectively and efficiently, if this is achieved by shifting burdens onto the very citizens those services are supposed to be supporting, how is that “better”? See Richard Pope’s report on Universal Credit[7] on whether the advantages of digitisation / automation are being shared fairly between DWP and UC claimants.

2.1 Boosting productivity and trade

It’s one thing to say data is hard to define; it’s quite another to misdefine it. Data is not knowledge, and the very assertion undermines what follows. Is this a National Data Strategy or a National Knowledge Strategy? (Noting the forthcoming “Digital Strategy, which we will be publishing in the Autumn”...) Definitions matter!

On that theme, if data is so fundamental to our economy then why talk in terms of a separate “data economy” or “data market”? What is the difference between “data-driven trade”, “digitally delivered trade”, trade that is facilitated by data (i.e. all trade), and trade in data itself?

It’s probably far more important to come up with clear definitions of each type of trade, and then develop methodologies to measure them, and then to provide some numbers, than to wave around a bunch of near-meaningless estimates. Certainly if one wishes to be able to measure the effects of one’s strategic interventions...

Data trusts are indeed “novel”. And until or unless they find a way to robustly manifest and enforce people’s human and other rights, they will remain a novelty. Just another type of ‘information intermediary’ that failed to get beyond early prototype.

It is notable that despite being a “world leader” – and despite its support for “institutions” such as CDEI, ATI and ODI – the first case study DCMS quotes is of a German packaging company sharing data with a Spanish logistics firm. That DCMS couldn’t find a UK example of third party Just-in-Time supply chain analysis doesn’t exactly bode well...[8]

2.2 Supporting new businesses and jobs

Fatima’s new job may indeed be in cyber. But that will likely have nothing to do with a National Data Strategy that not only doesn’t identify the data skills required, but fails to even indicatively map them across industry sectors, and remains mute on career progression / pathways, transferable skills, retraining, etc.

If you are going to draw a distinction between “data skills” and “digital skills”, then define them!

[7] https://pt2.works/reports/universal-credit-digital-welfare

[8] From his own previous work, the author could provide an example of this being done by a major British brand as far back as the 1990s.

Five years pre-COVID, much discussion was focussed on the loss of jobs due to automation.[9] Where is this thinking reflected in the NDS? Where's the case study on how Deutsche Bahn, the German Rail and Transport Union (EVG) and their workforce are navigating mass automation (2016-2023)?[10] How about medConfidential's proposition to (re)train postdocs on the Wellcome Trust Fellow model to do AI in the NHS?[11]

2.3 Increasing the efficiency and scope of scientific research

To which "five principles" on the "fair, ethical and appropriate use of health data" is the NDS referring? There are ten principles in DHSC's 'Code of conduct for data-driven health and care technology',[12] and five principles in DCMS' 'Data Ethics Framework'[13] – but the latter is for how data should be used across the whole public sector, not just health. And, as the NDS itself makes clear, DCMS is simply not competent to develop a Data Strategy for Health and Social Care. (Whether NHSX is itself competent remains to be seen...)

What evidence does DCMS have of "significant limitations on research" – other than the special pleadings of those lobbying for greater data access – and what *specific* "barriers to accessing data" are causing these? If any actual changes are to be made, it is going to be essential to distinguish between "real and perceived" legal barriers as the 'solution' in each case would be very different.

Claims of "delays and uncertainties" are as likely to be because companies' (initial) applications for data have been unlawful, unethical or unjustified. This is certainly the case with NHS patients' data, as evidenced by the number of malformed and inappropriate requests for data dealt with by IGARD at NHS Digital – not to mention the not insignificant effort of fulfilling some requests, e.g. where complex linkage across datasets is required.

The characterisation of the third paragraph is lazy and one-sided. Its final point, on the cost of the data itself, is complete nonsense given NHS Digital charges on a cost recovery basis only. Is it DCMS' strategy to give NHS patients' data away at a loss, or for free?

There is no "permissive approach" to data sharing during the pandemic; the Government awarded itself extraordinary emergency powers under the COPI Notices to take data that it would otherwise have no lawful basis to take. And it took that data for many purposes *other* than patients' care. If DCMS cannot even distinguish between data sharing with and without (citizens') permission, between mandatory and consensual, then its strategy is built on very shaky foundations indeed!

Maybe, rather than assuming just because government has been able to do some things during a national / global emergency that people's attitudes have changed, the architects of the strategy should actually read below the headlines of recent polling done by the National Data Guardian's

[9] The Bank of England's chief economist suggested of the order of 15 million jobs being "lost to robots" in 2015: https://www.theguardian.com/business/2015/nov/12/robots-threaten-low-paid-jobs-says-bank-of-england-chief-economist

[10] This being but one example of the ongoing and much-studied 'Industry 4.0' transformation taking place: https://onlinelibrary.wiley.com/doi/full/10.1111/irj.12291

[11] https://medconfidential.org/wp-content/uploads/2020/09/2020-09-vision-for-health-AIs.pdf

[12] https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology

[13] https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework-2020

Office[14] and others, and see (consistent with other research over years) that 1 in 4 people are likely to opt-out of uses beyond their own care[15] – and, in other polling, that 7 out of 10 want their 'COVID data' deleted at the end of the pandemic.[16]

Guess what? In a public health emergency, most medics and many professional medical researchers performed admirably. What is notable, though, is that those who have experience of handling patients' data at scale tended to perform much better; they maintained and even improved their information governance (given the imposition of the COPI Notices), and – *given adequate resources* – were able to accelerate the deployment of much-needed trustworthy infrastructure.

By contrast, public and commercial bodies with no experience of front line delivery of data (or to patients) have wasted enormous amounts of money, demanded 'all the data' but delivered little, and – despite promises – have remained untransparent about their ongoing use of patients' data, thereby corroding public trust.[17]

[General point: choice of case studies in NDS is pretty poor, sometimes to the point of irrelevance.]

2.4 Driving better delivery of policy and public services

It is undeniable that data can revolutionise the public sector, but broad-brush aspirations simply won't hack it. A proper strategy would learn from the decades-long digitisation of a sector (e.g. health) and where and how things succeeded (e.g. interoperable GP IT systems) and where and how they failed (e.g. massively centralised NPfIT, care.data, and 'data lakes').

This 'strategy' seems almost entirely focused on outcomes or end results – "better decision-making", "better, more coordinated use of data" – without (a) defining "better" in anything other than vague, general terms, and (b) with no articulation or indication of *how* these goals are to be achieved. What core principles need to be in place; what mechanisms of governance, oversight, incentivisation, enforcement, etc.

Before quoting an example as a 'success', it's probably worth checking that (as happened with the Shielded Patient List in its initial stages) the Government itself didn't utterly screw things up, having to be 'bailed out' by the experts. When speaking to examples throughout, it's best not to choose ones that demonstrate to domain experts that you really don't know what you are talking about. That the vaccine management process inside NHS England currently involves a 60m row spreadsheet suggests the highest priority data project of the last decade is likely to go badly.

While the public might possibly welcome more targeted support and services, they first expect that support and services will be funded and available. The first expectation of the public is delivery, not targeting – 'research' that asks narrowly limited questions often doesn't paint the full, or even an accurate picture. If you ask the public whether they expect the public services to track and target them like Facebook or Amazon, you will probably get very different answers...

[14] https://www.gov.uk/government/news/polling-indicates-growing-public-understanding-about-importance-of-using-health-and-care-data

[15] https://twitter.com/EinsteinsAttic/status/1316712864086986753

[16] Not an approach that medConfidential would necessarily recommend - but one that becomes almost inevitable if the data is not moved into a properly-governed COVID-19 Disease Registry, as we have proposed.

[17] Yes, we mean DHSC and NHSEngland(/Improvement) – otherwise known as NHSX.

Also, “targeting” in the public services is far more akin to ‘rationing’ than it is an online store. Service users of government are not ‘customers’ – they are citizens and taxpayers, and the expectations and obligations work very differently than in Silicon Valley or Redmond...

Again with “bureaucratic burdens”! What are they? Describe them, and provide actual examples. If what NDS is describing is unnecessary red tape then it can be removed. If, however, they are statutory safeguards like Data Protection, Human Rights, or Equality law then that’s a completely different matter. DCMS must say so, and follow the proper democratic path to new legislation in Parliament. (Noting the statutory basis for the COPI Notices will run out eventually...)

Data First is a nice try as a case study. And someone at DCMS obviously spoke to someone at ONS to know about the Five Safes. But where’s, e.g. the reference to Dr Byrom’s work for HMCTS, and how data used properly there can ensure that the Courts and Tribunals meet their statutory duties under the Equality Act? (That might fit better in the next section, but it was DCMS' choice to mention MoJ.)

2.5 Creating a fairer society for all

The incentives that have led to data ‘silos’ across government play out even more aggressively when you are forced to scrabble for funding (for survival). A strategy that seems not to appreciate or acknowledge the basic incentives of the entire Third sector is likely doomed to failure.

Assertions that AI is “helping tackle misinformation” when it is equally accurate to say that AI is being used to fuel the propagation of misinformation, not to mention the endemic bias of AIs trained on historic data sets, and the harms and dangers of online profiling and ad tech that is threatening the very democratic process itself.

“Biases arising from data or algorithm use will need to be addressed” is hardly a strategy; it merely (re)states the bleeding obvious. How will biases be addressed? It won’t be solely by data or technology – much as the data maximalists and AI evangelists would have you believe. The introduction of algorithms or ML/DL in the public sector is always a technosocial intervention.

DCMS would do best to understand this, develop an approach and rewrite this entire section. At present, it is close to meaningless drivel.

2.6 Realising the Data Opportunity

Congratulations! You’ve just noticed what happens when, rather than educating for understanding and core competencies, you teach a generation how to drive Word and Excel – or, half-heartedly, a bit of Python. If a part of your Data Strategy is not a serious rethink of (aspects of) formal education, ongoing lifelong learning and professional development then the UK will continue to be playing catch-up. And continue falling behind.

This isn’t just a “data skills” gap. Its about entry points and choices; genuine digital literacy – for parents, not just for kids; remember ‘On The Move’?[18] – which might also help with ‘online harms’

[18] https://www.youtube.com/watch?v=ufVe521quok

and misinformation; career development and transition; enforceable professional data ethics, rather than more endless lists; etc.

A-a-a-and we're back to the "Pillars" (see comments above).

3. Missions (see also comments on p5 above)

Mission 1: Unlocking the value of data across the economy

"Much of the transformative potential of data lies in the potential for linkage and re-use of datasets across organisations, domains and sectors" – does this mean data linkage and re-use across government, in the commercial arena, or of public sector (non-personal) data *into* commerce? Each of these three demand different approaches, and such lack of precision undermines the notion that this even is a "strategy".

If it proposes a "considered, evidence-based approach" why then does DCMS not unpack the sometimes wildly differing incentives to "collect" and "curate" data? In the public sector, data is by and large *collected* in order to deliver a service – i.e. its primary, legal purpose (such as health data to deliver health care) – but may need to be *curated* for other, secondary purposes that may be lawful, even desirable, but which are not why the data was collected in the first place. Especially if this curation requires changes to what data is collected or how it is collected, this is an additional (cost) burden on the service – so who pays?

The "balance" in secondary uses of personal data is not simply between "individual rights" and "public benefit" but, in many if not most instances, between individual rights, public trust and public benefit. Failing to factor in critical 'externalities' like trust is analogous to failing to attend to toxic polluting effects. (Noting that for government in particular, one data programme or initiative going wrong can have much wider toxic systemic effects.)

What happens when DCMS' research shows the quality of the data is much lower than it thinks? Layering crap data on top of crap data generally won't improve things – in what way does the strategy propose to deal with this?

Why the obsession with "short-term quick wins"? These may be politically or tactically desirable; they are often not how you achieve strategic aims. (Side note: exactly what "expertise" is CDEI supposed to have?)

"There are a number of ways the government can intervene to achieve this goal – including as a collaborator, steward, customer, provider, funder, **regulator** and **legislator**" – that DCMS does not understand why it cannot be both of these only strengthens the case for the ICO to be moved to MoJ...

Q5 is just ridiculous! Who is likely to answer in anything but their own special interests? Does DCMS propose to use answers to this question as evidence of 'demand' in different sectors? What independent research has it already done / commissioned? What types of data does each sector make use of, and what part (if any) can government play in each?

Q6 presumes central government *should* have a role; if this strategy is anything to go by, that presumption is rather questionable.

Q9 refers to “Smart Data”, which is a complete misnomer. Why is it “smart” for a person to allow a third party to share information about them with other third parties? Doing so simply perpetuates many of the problems within the current data ecosystem.
Government should instead lead by example and return information to individuals in re-usable form – as verified digital credentials, not just personal data – thereby empowering citizens as full actors in the data economy. Claiming to “put people’s data to work for them” is a loser argument if all your approach does is to perpetuate them as product or profit centres for others. Establishing the principle, the operation and the form of such credentials would be a truly strategic move; government achieves a far stronger mandate, and potentially far greater impact, where it makes *enabling* moves.

Mission 2: Securing a pro-growth and trusted data regime

Your priorities are showing! “So it is vital that the UK has a data regime that promotes growth and innovation for businesses of every size, while maintaining public trust” gets things the wrong way round. The only way the government can get this right is if it *first* ensures the UK has a data regime that maintains and assures public trust, while promoting growth and innovation for businesses of every size.

Why the split between the government's ‘Data’ and ‘Digital’ strategies? There may be a good reason, but it should at least be explained.

Bluff and bluster on “data adequacy”; what exactly will the government do? What changes to UK law will it make to ensure adequacy, which is something other sovereign entities get to decide? If it is as “vague” (DCMS’ own word!) as this strategy, then we’re in trouble...
The “widespread uptake of digital technologies” sounds pretty ‘digital’ – what data-specific measures are being proposed? What “compliance burdens” will be lifted? Examples?

Q11 & Q11a – DCMS is clearly a fan of its own creation, CDEI, and would equally clearly love for its stooge to be put on a statutory footing. This would be a terrible idea. (Not least because its chair is implicated in several of the worst data debacles in recent UK history[19]). Beginning Q11a with the word “How” is effectively push-polling; a better formulation would be: “Would a change to statutory status support the CDEI to deliver its remit?”

Mission 3: Transforming government’s use of data to drive efficiency and improve public services

The “high watermark” of data sharing to which DCMS refers is under emergency powers during a global pandemic! It is unrealistic to expect such powers or conditions to persist, and to pretend that the outcomes of all such data sharing have been positive is a gross misrepresentation. It’s all very

[19] https://news.sky.com/story/chair-of-embattled-exam-watchdog-owned-data-firm-that-was-involved-in-major-nhs-care-scandal-12054707

well to talk about improving delivery and measuring impact, but absent *publication* how are the public to discern these effects?

DCMS proposes to have a "Government Chief Data Officer", despite having failed to recruit one over three years after the role was first announced.[20] Meanwhile, Cabinet Office is recruiting a Government Chief Digital Officer[21] which – along with the split between the Data and Digital strategies themselves, and the assertion that "We need to transform the way data is collected, managed, used and shared across government" – does rather beg the question of how 'joined up' this all is. The prospect of yet more Whitehall power games and territorial pissings fills many who actually care about these issues with dread.

> ***Quality, availability and access****: striving towards improved data quality that is consistent, a clear understanding of what data is held and where, better data collection, and efficient data-sharing between organisations. All should be the norm, rather than the exception.*

While one can aspire or even aim for a consistent *level* of quality, different types of data are inherently different. Mapping where data is held is a necessary step; it may not be where the data *should* be held, though! "Better" always begs the question "better for who?" – the public, the public services, or for secondary users / commercial exploitation? Data *sharing* is not always the most efficient way to deal with data (creating a proliferation of copies, encouraging siloing, etc.) and the clue is in the word "access" – cf. safe settings, especially for sensitive data.

> ***Capability, leadership and culture****: developing world-leading capability in data and data science across central and local government, so that leaders understand its role, expert resource is widely available, staff at all levels have the skills they need, and a 'data-sharing by default' approach across government tackles the culture of risk aversion around data use and sharing.*

You're going to have to pay more, then! Unless you hadn't noticed those skills you're talking about are much in demand in the commercial sector too. And if the "culture" you are trying to embed is "data sharing by default" – rather than, e.g. canonical registers, safe settings, etc. – than your strategy is truly screwed (and hasn't changed much in the past 20 years).

> ***Accountability and productivity****: opening government up to greater scrutiny and increasing accountability, ensuring that this drives improvements in productivity, policy and services for people, while also ensuring data security; and using procurement to drive innovation and better outcomes.*

Accountability is good. We'll believe it when we see it.

If we're talking about "high watermarks" of the pandemic, then some of the ring mark stains to which government must attend are the appalling lack of transparency around data use, e.g. of NHS

[20] https://www.publictechnology.net/articles/news/government-chief-data-officer-job-remains-unfilled
[21] https://www.publictechnology.net/articles/news/recruitment-begins-government-chief-digital-officer

England's COVID-19 Data Store,[22] Palantir's Foundry platform,[23] as well as by other commercial entities – Faculty AI,[24] Deloittes,[25] et al. – and the persistent deviation from established procurement frameworks and due process that cannot help but whiff of corruption. Also that the very public evidence is that such "innovation" has in no way led to "better outcomes", cf. Test and Trace, etc.

When many very recent, high-profile, eminently citable counterexamples exist, DCMS would be wise to either tone down or provide hard evidence to back up statements like this:

> ***Ethics and public trust****: this transformation will only be possible and sustainable if it is developed within a robust ethical framework of transparency, safeguards and assurance which builds and maintains public trust in the government's use of data.*

You missed "legal", as in "robust **legal** and ethical framework". What Government does is laws; when it comes to ethics, not so much. A starting point might therefore be for government to ensure that all of its systems and delivery complies with, respects and enhances human rights, data protection, freedom of information, equality and all other relevant laws. And to not introduce any legislation that undermines or diminishes any of them.

Mission 4: Ensuring the security and resilience of the infrastructure on which data relies

["Infrastructure" in this context seems more 'digital' than 'data', but hey...]

Does the line, "we will also take a greater responsibility in ensuring that data is sufficiently protected when in transit" mean DCMS will robustly oppose moves in some parts of Government to undermine end-to-end encryption? If not, that will undermine all the good that the strategy says comes from data.

Has DCMS considered the extent to which a "Cloud first" policy, or 'outsourcing' the processing of bulk data to, e.g. US providers, exposes citizens' data to increased risk? One would hope the government already has its own good answers to Q14a – and that they are embedded in procurement frameworks like G-Cloud and Digital Marketplace. The extent to which the government can assist, e.g. SMEs to achieve the same (or better) levels of assurance is an open question. Will, e.g. GCHQ/NCSC be getting into the business of 'kite marking' cloud-based offers?

For the most sensitive types of data, e.g. genomic data, has an assessment been done of the required investment to protect what is (and will increasingly become) critical national data assets? There's little point trying to position the UK as a 'life sciences economy' if the intention is simply to hand over copies of the data to Silicon Valley outfits, which get to exploit the derived insights...

[22] http://usemydata.org/resources/use%20MY%20data_Webinar%200902920_Letter%20to%20NHSX.pdf This, despite public promises in March: https://healthtech.blog.gov.uk/2020/03/28/the-power-of-data-in-a-pandemic/, civil society requests in May: https://www.theregister.com/2020/05/19/covid19_nhs_data_store_open_letter/ and - having exhausted all other routes - our own letter to the CMO in June: https://medconfidential.org/wp-content/uploads/2020/06/2020-06-01-medConfidential-letter-to-CMO.pdf

[23] e.g. https://tech.newstatesman.com/coronavirus/palantir-nhs-datastore-contract-extension

[24] Whose involvement goes far beyond health: https://www.telegraph.co.uk/technology/2020/08/24/vote-leave-ai-firm-handed-new-government-contract-map-covid/

[25] https://news.sky.com/story/coronavirus-more-than-1-000-consultants-from-deloitte-on-test-and-trace-programme-12099127

Q17 – the best way for Government to ensure “that data use does not negatively contribute to carbon usage” is for it to clean up the UK’s energy supply. And this doesn’t mean simply offshoring the carbon consumption; measures taken should contribute to a global net reduction.

Mission 5: Championing the international flow of data

Having left the European Union, the UK has a unique opportunity to screw things up by undermining existing data standards such as GDPR and participating in – if not actively provoking – a ‘race to the bottom’. Seeking ‘adequacy’ with legislative regimes that lack any sort of general data protections (such as the US) would be a fatal mistake. If anything, the UK should be championing international rules that are GDPR+, not GDPR-.

It is unclear to what extent the UK will be able to *enforce* the “regimes, approaches and tools” it adopts internationally. It is one thing to do so as part of the GDPR ‘bloc’; to try to do so as an independent ‘outlier’ that is itself seeking to maintain adequacy is an order of magnitude more complex. There is little indication that the UK by itself has the power or is even competent to regulate major international data corporations – except through what amounts to trade boycotts, as with 5G.

What “ambitious data provisions” does DCMS propose? Detail matters! Focusing on data transfer alone is fatal in a trade and economic context, where issues such as Intellectual Property are equally if not more relevant. Data itself is more like a raw material in this context; the real value resides in the insights and inferences, and the services and products developed. Merely focusing on the flow of data as if this alone is sufficient is far too narrow a view.

> ***Facilitate cross-border data flows**: We will work globally to remove unnecessary barriers to international data flows. We will agree ambitious data provisions in our trade negotiations and use our newly independent seat in the World Trade Organisation to influence trade rules for data for the better. We will remove obstacles to international data transfers which support growth and innovation, **including by developing a new UK capability that delivers new and innovative mechanisms for international data transfers.** We will also work with partners in the G20 to create interoperability between national data regimes to minimise friction when transferring data between different countries.*

What exactly is this “new UK capability”? What “new mechanisms” will it deliver, and how are they “innovative”? Making broad-brush assertions like this without any indication of what they mean in practice – or even in principle – is not strategy, it’s wishful thinking. Or bullshit. If DCMS has any sort of proposition, it should publish it.

There is much talk of “incorporating”, “driving” and “promoting” UK values (and ethics) through data. It is far from clear, however, that UK values of “openness, transparency and innovation” are (a) values unique to the UK, (b) values to which the UK successfully holds itself, and (c) a sufficient set of values to achieve the intended goals.

If, for example, the UK wishes to “build trust in the use of data” than it must ensure that what it does with data is *trustworthy* – i.e. reliable, competent and honest (as well as sharing a common moral / ethical framework) – of which openness and transparency are but components. Not to mention the fact that the UK Government's performance with regard to transparency to its own citizens has fallen off markedly[26] over time.

If the UK seeks data adequacy with the EEA, then it must not only maintain step with GDPR but sort out the existing issues that already threaten adequacy. Without explaining how it will do this, Q18 and Q19 are basically pointless; just because you *want* something that meets two different sets of different criteria does not mean that there is a solution that fits. DCMS must publish its proposals for people to assess / comment on.

[26] https://www.instituteforgovernment.org.uk/publication/whitehall-monitor-2020/transparency

4 Data foundations: ensuring data is fit for purpose

> *In this strategy we are using the term 'data foundations' 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 basically, as you say in the 5th paragraph of the introductory section, you just want to *"pool data from multiple sources and sectors"* – i.e. create yet another 'data lake'. This is at best last-generation thinking, and largely discredited as an approach.

And again, what precisely does DCMS mean by "fit for purpose"? If by "advanced technologies" it means "AI" (machine learning / deep learning) then, in order to be fit for purpose, the data will often not only need collecting and collating but labelling. Who does DCMS propose should do this? And why should this one purpose be prioritised over investment in other more basic and likely more widely applicable processing?

In its obsession with "innovation", and when shilling for its AI mates, CDEI tends to miss out on a whole range of more mundane but useful opportunities and benefits.

4.1 Data foundations in the wider economy and society

Government can encourage adoption of standards in a number of ways – through its own procurement, and by mandating them in laws or regulations (though the latter may impose actual burdens, not just perceived ones). A strategy that says "standards are good" is really not saying much, and clearly "one size will not fit all" – but you would expect a strategy to outline an approach or approaches to the problem, rather than just name it.

On ADR UK and the Digital Economy Act, see our mid-term review.[27] It'd probably be wise to demonstrate some actual benefits before "extending this approach to other areas of the economy" – especially when those "other areas" seem to be predominantly health data.

Businesses use and do what works. Regardless of Government research or evidence, they are unlikely to make different "resourcing choices" until or unless there are clear and readily available benefits to their own bottom lines. Unless their business actually involves serving the government in some way, or meeting Government-imposed statutory or regulatory requirements, making data fit for government purposes is unlikely to be a priority.

On 'legacy lock-in', the government would do well to take a long, hard look at itself first! If it intends doing research in this area, government itself will provide many examples of the problem. Whether it provides many good examples of how to tackle or fix it is less certain.

Does the new "UK R&D roadmap" include health data? If so, why does that form part of this strategy and how will it align with the as-yet-unwritten Data Strategy for Health and Social Care?

[27] https://medconfidential.org/wp-content/uploads/2020/06/medconfidential-Digital-Economy-Act-Review-note.pdf

Rather than “bolstering efforts to ensure that consumers’ data is put to work for them” how about making sure that citizens can see how their data is used, and giving them meaningful choices about that? Simply taking people’s data and “making it work for them” without choice / consent and transparency risks imposing precisely the sort of ‘one size fits all’ approaches that DCMS says it wants to avoid. And it would likely be corrosive to public and individual trust too.

4.2 Data foundations across government and the wider public sector

That something as apparently straightforward as address fields are not treated consistently across government suggests DCMS may be being a bit optimistic (though it’s good to have ambition!) about standardisation and interoperability. This is an area where a central or overarching approach could be beneficial – cutting across inter-Departmental power games – but it seems unlikely that DCMS has the ‘clout’ to be able to do this, when even Cabinet Office struggles.

And, in practice, different parts of the wider public sector are most likely to orient to their relevant Department. While DCMS might want to drive better “data foundations” in, say, health it is unlikely to do so without the agreement and assistance of DHSC. (It is interesting that a Data Strategy which explicitly defers to a separate Data Strategy for Health and Social Care keeps returning to health examples. Can DCMS not find examples from other areas?)

The word “integrated” appears 8 times in the March 2020 Budget,[28] but none of them make mention of “the development of an integrated platform for data across government”. It appears ONS will be “leading the effort to bring an Integrated Data Platform (IDP) for government”[29] but very little detail is forthcoming. Indeed, we are finding out more about it from ONS’s business plans (its 2020-25 deliverables timeline says it will “Deliver the Full Business Case for the Integrated Data Platform Programme” by March 2021,[30] while acknowledging the need for HMT budgetary sign-off) and from job advertisements[31] (such as quoted immediately below) than from government itself:

> ***About The Integrated Data Platform (IDP)***
>
> *The IDP will create a safe, secure and trusted infrastructure for Government data, enabling analysis to support economic growth, better public services, and improving the lives of citizens. It will do this by:*
>
> - *Building on ONS’s existing infrastructure to ensure analysis is carried out (where legally and ethically appropriate) across Government in line with priority needs; with data being made available to the research and academic community to support this analysis.*
> - *Expanding the analytical capacity and capability in Government for analysing large-scale, linked datasets.*
> - *Ensuring findings and analysis are accessible and used.*
> - *Facilitating better evaluation of the effectiveness of government policy.*
> - *Ensuring common standards for bringing analytical data together, working alongside Government Digital Service (GDS) and Department for Digital, Culture, Media & Sport (DCMS).*

[28] https://www.gov.uk/government/publications/budget-2020-documents/budget-2020
[29] https://www.ons.gov.uk/news/news/nationaldatastrategytheonstakescentrestage
[30] https://uksa.statisticsauthority.gov.uk/wp-content/uploads/2020/07/ONS-Business-Plan-2020.pdf
[31] e.g. https://www.uxresearchjobs.com/job/user-researcher-881583

- *Leading the government's ambition for a step change in data capability across government*

This is an exciting time to join this complex programme which is being mobilised at pace.

Such an initiative, that will clearly be processing vast quantities of citizens' data for a whole range of purposes, demands impeccable transparency from the outset. Start as you mean to go on. And don't even think of trying to hide behind "anonymised" this time!

We would hope – but note the lack of any form of confirmation – that as an ONS project, what will be produced is publishable analytical and statistical outputs available to all, not more policy-based-evidence-making. Any belief that a patients' medical records could be used by ONS / HMG to justify closing their local hospital, is likely to be toxic – not just in itself, but to ONS, Official / National statistics, the data analytical infrastructure, and *bona fide* research uses of data.

Experience has shown that while "a central team of experts" may be able to "ensure a *consistent interpretation* of the legal regime around data sharing", that doesn't mean that the interpretation itself will be legal. Indeed, some of the worst data scandals (e.g. care.data and abuse of HES data) arose precisely because of 'eccentric' interpretations of the legal regime. Competence and expertise needs to be embedded in the organisations who are doing the sharing, and who have intimate knowledge of the data itself – anything more remote is simply asking for trouble.

Also, once again, things are focused altogether too narrowly on 'data' itself. Whither Data Protection, duties of confidentiality, or Information Governance, for example? What about the legal requirement for DPOs – a statutory role in primary legislation sponsored by DCMS that is not mentioned *once* in the entire strategy? What about Child Protection and Safeguarding Governors in schools, or equality obligations on employers and in HR? DCMS should take a good look at what is already in place before 'wading in' with its own team of "experts", who are likely to be anything but expert in every sector...

Phrases like "government's own data" tend to make people nervous. To what data is DCMS referring in this instance? Is it clear that none of it is citizens' personal data? If it is, then why call it "government's own"?

Being a data controller for personal data does not give one carte blanche to do with it as one wishes. Indeed, a more appropriate analogy might be that of a data guardian or custodian – looking after something valuable on behalf of those to whom it relates. If your answer to removing "technical barriers to data use and re-use across government" is the creation of a central government data lake, then trouble is virtually guaranteed.

How will your "spend controls" deal with off-framework procurements that use Excel for critical data transfers, or produce "mutant algorithms"? Mandating standards does not deliver assurance; they must be properly audited and enforced. There seems to be a degree of magic thinking here: "If we tell people to use standards they will all fall in line". Does DCMS seriously expect HMT or HMRC to bow to it on data? Or DWP or the NHS, for that matter?

A lot seems to ride on DCMS' new Chief Data Officer, whose relationship to the Cabinet Office's Chief Digital Officer are not yet finalised, and on as-yet-unspecified powers / authority to enforce standards and effectively to perform audits (e.g. "review of governance structures", "ensuring central government departments include data management plans") across government.

Given DCMS just lost a chunk of digital policy to the Cabinet Office, how realistic is this? And where are the draft proposals for these powers, which would clearly require primary legislation if they are to have the stated effect.

4.3 Supporting data foundations internationally

Unsurprisingly, this is one of the sketchiest sections in the document: “We want to be a data champion across the world” can hardly be described as a ‘strategy’! There is plenty of talk of “partners”, “like-minded states”, and shared “values” – but what of those states that *don’t* share our values, or even a common understanding of human rights? This reads more as if DCMS is playing checkers with an 8 year-old friend, than playing the sort of 4-dimensional chess that takes place on the global stage.

What does DCMS propose to do when a country is resistant, or actively hostile, to our “values” or data demands? Any competent strategy would at least consider and make suggestions as to how it would address such eventualities. (This is but one example of where the strategy simply omits to mention difficult problems. One has to be positive, but also realistic.)

5. Skills: Data skills for a data-driven economy and data-rich lives

Given “There is no widely agreed definition of data skills”, that DCMS has commissioned research to produce a definition is a good thing. Having a ‘taxonomy’ of skills (data vs digital vs AI, etc.) may also be useful, but by far the most helpful thing will be mapping these to their real world use, e.g. to practical pathways into, through and across careers – and not necessarily just “data related” ones.

If this is to be a truly national strategy – underpinning the success of the UK economy, citizens’ participation in it, and lack of exclusion from it – it is vital that these are not just seen as desirable “skills”, but necessary *competencies*. To achieve the broad base required to support an engaged citizenry and a competitive data economy in the global information society, what we should be talking about is something more akin to basic literacy or numeracy.

We already know there are many data jobs ‘further down the stack’ which are currently outsourced or delivered using mTurk. When AI replaces humans in the call centres, what is the equivalent-level job for those who will have been displaced? While the aspiration must of course be that there are pathways to ‘a good job’ for everyone, the reality is that every economy rests upon a layer of jobs that are (or must be made) just ‘good enough’.

5.1 Driving clarity and coordination

The proto-institutions mentioned are all relatively new, and at least some of the “confusion” arises from the fact that they are still jostling for position in an uncertain and rapidly-changing territory. If DCMS wants to promote professionalisation – which, not coincidentally, can be a critical factor in the embedding and enforcement of ethics – then it would do well to look at well-established models (e.g. in medicine) to see how they evolved, and are evolving[32] to deal with the emergence of data science and informatics as fields of practice.

5.2 Ensuring formal and vocational education rises to the challenge

The ‘simplification’ of post-16 vocational education with T-Levels[33] is a bold experiment, but if it is only the “technical qualifications” that include data / digital skills, how will this address general digital literacy and inclusion?

The Airbus example is an interesting case in point – a multinational company that clearly sees a competitive advantage in investing in and improving the general data capabilities of its multinational workforce, regardless of where they were educated.

The ‘Office for Talent’ is a measure necessary to counter the hostile environment that still operates in the UK, the effects of which will if anything only become worse after Brexit. Such a move also presumes that people will want to come to work in the UK. Maybe the strategy should attend to the equally likely prospect of a ‘digital brain drain’?

5.3 Driving data skills across the public sector: capability, leadership and culture

This first sentence from the second paragraph of this section is a pretty damning assessment...

> ***The lack of a mature data culture across government and the wider public sector stems from a fragmentation of leadership and a lack of depth in data skills at all levels.*** *The resulting overemphasis on the challenges and risks of misusing data has driven a chronic underuse of data and a woeful lack of understanding of its value.*

...but it is not clear that the analysis in the second sentence necessarily follows from it.

While there is indeed a “woeful lack of understanding” of data and digital across large parts of government, this does not seem to translate into an underassessment of value, if officials’ incessant demands for “More data!” are anything to go by. And any “overemphasis on the challenges and risks” does not seem to stop Ministers and Departments from attempting stupid and dangerous things with data, time and time and time again. (It is also difficult to determine “chronic underuse” when the government often cannot be clear about what its genuinely useful uses of data are or have been...)

Would these innovation “Fellows”, sponsored by No.10, happen to be part of Faculty, perchance? And what transparency will there be on the direct involvement of “the digital and tech sector” in “transformation projects of national importance”?

In line with previous comments, and rather than focusing purely on “data expertise” across government – and given the excoriating self-criticism that leads this section – would it not be sensible to focus on data *competency* as a goal? Those who are (or become) expert may be drawn to data-related or data-specific functions, but it is the government's *general* failing to understand data and digital that leads to its poor data culture, and the chronic mismatch between policy and delivery.

[32] https://facultyofclinicalinformatics.org.uk/
[33] https://www.tlevels.gov.uk/

Getting ONS involved is a good thing. But not if it degrades or devalues (official) statistics. The whole strategy is rather ‘light’ on its consideration of how greater use could be made of statistics in general – an approach that, done properly,[34] might also address some of the ‘fears’ around misuse of individual-level data.

6. Availability: ensuring data is appropriately accessible

Government refers incessantly to “unnecessary barriers” to data sharing and data (re)use. How do DCMS and other parts of government define the difference between a necessary and an unnecessary “barrier”? The non-comprehensive list of “barriers” given is illuminating:

- *a culture of risk aversion*
- *issues with current licensing regulations*
- *market barriers to greater re-use, including data hoarding and differential market power*
- *inconsistent formatting of public sector data*
- *issues pertaining to the discoverability of data*
- *privacy and security concerns*
- *the benefits relating to increased data sharing not always being felt by the organisation incurring the cost of collection and maintenance*

given they cover a spectrum of cultural, legal and regulatory, practical / operational, and financial / economic concerns. (And it must be noted that privacy and security are not “concerns” but necessary *preconditions* for lawful, appropriate data access or sharing. People may ‘hide behind’ these as excuses for not sharing data, but they are not in themselves barriers – unless what you are attempting to do is clearly wrong.)

On benefits, it is not just that the organisation incurring the cost of collection and maintenance may not feel them. Often, it is the people whose data is being shared who feel no benefit either. Until government ‘opens a channel’[35] from Departments and public bodies to citizens, and uses it to evidence the wider benefits and, e.g. engage citizens in the research endeavour, government and commercial data (re)use in the public mind will just keep delivering bad news.

Also, definitions matter. **Data access** is not the same as data **portability**, data **mobility** or even data **sharing** – and the terms should not be confused. The critical difference is where *copies are made* and *passed beyond the direct control of the data controller*. Secure research environments / safe settings and attribute or credential-based models ensure (personal) data transfer is kept to an absolute minimum; portability, mobility and sharing all involve the creation of copies of the data, which is where most of the problems start.

A strategy that is not absolutely clear on this point has not grasped one of the fundamental challenges of data (and digital). This is evidenced by the ‘scattergun’ examples adduced, ranging

[34] Individual-level data is rarely needed to inform policymaking; instead, define the questions that need to be answered and measurements that need to be made and develop appropriate statistics that can be published. Policies can be modelled, delivery tracked and evaluated, and improvements made over time – all in a transparent and trustworthy way.

[35] We proposed data usage reports, first in 2014: https://medconfidential.org/2014/what-is-a-data-usage-report/ and in more practical detail a year later: https://medconfidential.org/2015/implementing-data-usage-reports/

from yet more health data examples to sectoral frameworks, to unproven ‘prototyping’ like data trusts.

On the “balance between individual rights and public benefit”, acknowledging Article 8 of HRA 1998 permits a lawful, necessary (and proportionate) balancing with “economic well-being”:

> *There shall be no interference by a public authority with the exercise of this right except such as is* ***in accordance with the law*** *and is* ***necessary in a democratic society*** *in the interests of national security, public safety or* ***the economic well-being of the country****, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.*[36]

It would be helpful for DCMS and the NDS to acknowledge that data use engages laws other than the Data Protection Act 2018 and GDPR. The Equality Act 2010 being a case in point, especially in the non-discriminatory delivery of public services – but also many others, including the common law (e.g. common law duties of care and confidentiality in health).

6.1 Data availability for the economy and society

While Open Banking is a qualified success,[37] “Smart Data” is far from a panacea.

Interoperability and portability are certainly beneficial, and competition in a marketplace is generally good for consumers – until, e.g. someone comes up with an intermediary platform, or one ‘gatekeeper’ predominates. If the goal is to have third parties passing your data around between themselves, then the ambition is not ‘smart’. Indeed, it’s not that much different – or innovative – than what already pertains.

While pithy, it is not clear what getting “consumers’ data [to] work for them” actually means. Do consumers get a choice in this? Who is actually in control? Such a notion could manifest in several ways, not all of them good. As we have seen, companies can become incredibly successful (and powerful) ‘sweating the asset’ of other people’s personal data – not always with positive results for those people, or for wider society.

On public sector open data, I will simply note the tail-off of updates from most Departments to data.gov.uk over the past five years – most of which cannot even be bothered to keep their own organograms up to date, and many of which have a parlous record on FOI requests too.

If the government wants to lead on open data, it must lead by example!

On the energy sector and regulatory frameworks case study: what are “agile regulatory approaches”, and in what circumstances are these applicable? For example, while it is conceivable that a regulator could regularly monitor the performance of a limited number of regulated suppliers to the Grid, this would be very different to the situations that pertain around, say, personal data in advertising markets...

Note: “...the government is launching a £2.6m programme to help companies to develop AI-based solutions to tackle [online harms] issues ever more effectively.”

[36] https://www.legislation.gov.uk/ukpga/1998/42/schedule/1/part/I/chapter/7

[37] Claims of an average £12,000 annual benefit per user seem... optimistic. And such benefits are unlikely to be repeated year on year.

There’s a bunch of ‘cool’ initiatives here. Where’s the strategy?

6.2 Data availability within government and the public sector

The move from legal gateways for cross-government data sharing to the ‘framework’-based approach of the Digital Economy Act 2017 was described as absolutely “necessary” during the passing of the Act. Why then have these “necessary” powers barely been used?

For more on this topic, see our mid-point review of the Digital Economy Act, which we could cut and paste into here, but this is getting long already...

https://medconfidential.org/2020/data-misuse-as-missed-use/

6.3 International data availability

Not all data flows are equal. And there are some that you will wish to prevent.

No-one will thank you for copying critical national data sets (e.g. health data, genomic data, business microdata) to regimes with lower safeguards and protections and, in so doing, haemorrhaging public and economic value.

It seems that around trade, the definition of “data” becomes ever more imprecise. And with regard to value – where other concerns, e.g. derived insights, trained models, intellectual property, business models, etc. are engaged – a narrow focus on just the data or data flows risks missing critical aspects.

Will the “independent HMG capability to conduct the UK’s own data adequacy assessments” be separate or distinct from the ICO? If so, what is the rationale for that split? And, given the ICO’s degraded performance under DCMS sponsorship, is DCMS really the place where such decisions should be being made?

To what “alternative transfer mechanisms” is DCMS referring? Are they akin to the border that won’t exist in the Irish Sea? Or to ~~Safe Harbour...~~ ~~Privacy Shield~~... or whatever’s next?

7. Responsibility: driving safe and trusted use of data

> *In this strategy, we use **‘responsible data’** to mean data that is handled in a way that is lawful, secure, fair, ethical, sustainable and accountable, while also supporting innovation and research.*

At least they got the “while also...” round the right way this time!

Given “Responsible Data” is the subtitle of one of the ‘Pillars’ of NDS, the fact that the term appears just 9 times in the entire document – 3 of those as such subtitles, 4 of them in (verbatim repeated) definitions of the term, and zero times outside of this section (except as subtitles) –

suggests that “driving safe and trusted use of data” is maybe not the underpinning foundation to the strategy that it should be.

Couldn’t agree more with “The government must also be transparent and prepared to open itself up to scrutiny over its own use of data” – government must lead by example. And it is right to acknowledge that a legal framework *by itself* will not and cannot “earn people’s trust”; too many see mere compliance with Data Protection law as being sufficient.

Don’t cherry-pick your polling![38] The numbers are pretty stable, and have been for decades: *if people’s permission is sought*, a majority are content for their medical records to be used for (public good) research. *If their permission is not sought*, or they are not told what their data is being used for, a majority are not content. Perceived commercial exploitation is also a factor – as the ability for people to make an “active decision” around use of their data; a decision that is respected by all.

Also, it’s not just ‘transparency about benefits’. Tim tried that with care.data! It’s about honesty across the piece: people must know how their data is used, and be able to make meaningful choices about that.

So data (re)use by different actors is variously described, in this section alone, as having to be: “lawful, secure, fair, ethical, sustainable and accountable”; “lawful, secure, unbiased and explainable”; “collecting, storing and using [citizens’] data safely and securely, in accordance with the highest standards of ethics, privacy and security”; and with “fairness, transparency and trustworthiness”. And thus some common characteristics emerge:

- Lawful
- Safe and Secure (noting the difference between the two)
- Fair and Unbiased (noting the difference between the two)
- Ethical
- Transparent and Accountable (noting the difference between the two), also Explainable
- Maintaining the highest Standards of Privacy (i.e. beyond mere legal compliance)
- Sustainable (for who?)
- Trustworthy (i.e. Reliable, Competent and Honest)

7.1 A pro-growth data rights regime

If DCMS thinks the ICO’s sandbox is the way forward, then it should rethink! Giving a legal ‘free pass’ to any initiative that manages to get through the process (a company’s first act when the ICO tries enforcing against it for something it did in the sandbox is simply to JR the ICO...) and encouraging organisations to ‘push the boundaries’ of a law that is already being poorly enforced is not in line with any of the characteristics listed above.

> *A wide range of research suggests transparency around how data is used is important for building public trust,[39] and the importance of trust as an enabler for public sector data sharing.[40]*

[38] ONS, Feb 2020: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/903239/Feburary_2020_Opinions_and_lifestyle_survey_data_module.xlsx

[39] https://livingwithdata.org/project/wp-content/uploads/2020/05/living-with-data-2020-review-of-existing-research.pdf

[40] https://www.gov.uk/government/publications/cdei-publishes-its-first-report-on-public-sector-data-sharing/

No shit! So why are the government and public bodies not *doing* it? Commissioning endless studies, and issuing guidance for the latest shiny objects are no substitute for doing the basic work of publishing evidence of functioning IG and oversight processes, release registers, and preferably data usage reports (or 'statements').

"We will run a national engagement campaign on the societal benefits of the use of government data" sounds awfully like the same 'selling the benefits' approach of care.data. Rather than spending taxpayers' money on this, why not first put in place meaningful consent / dissent and transparency measures and then invest in properly promoting those? The former is just more marketing; the latter is far more likely to engender sustainable trust.

More algorithmic transparency would be good – but equally, if not more importantly, better governance and accountability around "algorithmic assisted decision making" of all kinds (not just "AI") is needed.

Privacy enhancing technologies are great, when they are implemented properly – but they are still only a 'technical fix'. Implementing meaningful choices and feedback mechanisms are ways to provide far greater and more obvious citizen control.

A "Data Ethics Framework"[41] that, e.g. encourages people to self-assess their compliance with the law on a scale of 0 to 5, is unlikely to have the desired impact. It's yet another compliance checklist. The National Statistician's self-assessment tool[42] – grounded as it is within an actual profession, i.e. research – offers a better model. And proper use and publication of tools like DPIAs could help identify good (and bad) practice.

The reason why the role of National Data Guardian for Health and (Adult) Social Care had to be created in 2014 was because of the care.data and HES scandals – and also because the equivalent role (chair of the National Information Governance Board) had been abolished in 2012.

There is really no equivalence between NDG and CDEI, and no evidence or apparent good reason to put the latter onto a statutory footing. To what or to who would its statutory 'loyalties' be? And in terms of ethics, it will need at the very least a change of personnel, given its Teflon chair has been implicated in several of the most high-profile data debacles of the past decade or so...[43]

7.2 Data use that is secure and sustainable

As with our pandemic preparedness, the WannaCry incident demonstrated that government and public bodies had been all too slow in investing in and addressing systemic vulnerabilities. While it can mitigate some risks, a move to the Cloud does not eliminate all risk – and indeed, poor implementation of Cloud technologies (such as ElasticSearch) can lead to far greater exposure of

addressing-trust-in-public-sector-data-use

[41] https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework-2020

[42] https://uksa.statisticsauthority.gov.uk/about-the-authority/committees/national-statisticians-data-ethics-advisory-committee/ethics-self-assessment-tool/

[43] https://news.sky.com/story/chair-of-embattled-exam-watchdog-owned-data-firm-that-was-involved-in-major-nhs-care-scandal-12054707

vast quantities of citizens’ data, not to mention mass surveillance of bulk datasets. A National Data Strategy must attend to security from bad actors ‘within’, as well as those ‘without’.

Ooh! Another strategy: “Greening Government: ICT & Digital Services Strategy 2020-2025”...

-ends-