

Draft: 24/10/2017

Available next steps in Data Policy for Government: Towards Data Flow Controls

AI has finally got institutions to take data seriously - especially in Government. Using the coincident timing of GDPR, the update to the Cabinet Office data sharing ethics framework, and **mirroring the “[spend controls](#)” for technology, there should be “flow controls for data”** based on information in publishable registers (as proposed in the consultations for the Codes of Practice for Part 5 of the Digital Economy Act).

Current data practice in Government is 5 different scandals away from adequacy. As our work in the NHS shows, some of those scandals will be public, some of those will be private - all are optional, and the improvements are inevitable. The range of discussions currently ongoing give the [opportunity](#) for the data choices of the future to be better than those of the past; but only if there's a way for the public to have an evidence based discussion. GDPR makes a substantive policy opportunity out of a process requirement. That opportunity includes data transformation and AI, but also whatever comes after digital transformation and AI.

Below is what that could look like, **using today's laws for tomorrow's benefits.**

What could Cabinet Office flow controls for data look like?

Taken together, data flows that are **grounded in public knowledge, and which cite the law**, will incentivise **better decisions and fewer scandals**, and help **increase the measurability and effectiveness of other reforms**. The goal should be allow a recipient of public services to be an informed participant, rather than a passive observer.

The National Statistician's Data Ethics Committee is a good start, by addressing the highest profile and precedent setting issues across Government. However, as with other parts of the Digital Economy Act (Part 5), there should be a Data Review Board (equivalent) for all sharing agreements. It gives a process for which data sharing decisions can be reviewed, retroactively if needed (based on new information). However, if the only evidence any panel hears is from a scheme's designers and proponents, then of course care.data will be signed off as a good idea by those who are given a business case without the context that they may wish to have...

The [replacement](#) for the data science ethics framework should become a key overview document, describing the data to be copied, the law which allows the copying, and also what the citizens whose data will be copied will be told. This should be part of documents that the GDPR and policy already requires. In effect, that framework should become the overview for the Register, which covers how any data has been used. And on what basis.

Citizen facing departments can also use their digital services to inform their citizenry of how each individual's data has been used. The [Spring 2017 Conservative Manifesto](#) gave a strong and clear vision of how Gov.UK Verify would be at the heart of a Government that was accountable to its citizens,¹ and the benefits of that vision apply here. When there is a data debacle, that individual channel to each citizen allows for targeted information.

¹ page 3: <http://medconfidential.org/wp-content/uploads/2017/09/government-supplement.pdf>

1. Registers: public knowledge about data flows

Flow controls would ensure that all flows of personal data between Departments are listed in a public register, no matter their legal gateway. This would allow a shared understanding of where data is not shared but should be, or vice versa. As a matter of good governance, both failures are important, although only one tends to receive significant attention.

Ministerial intent is not necessarily what gets implemented in practice on the front line of delivery. While modern media provides a feedback loop from the citizen to the Minister's office, finding out facts is often hard for all. Citizens are not provided with a factual basis of what data is or should be shared. Departmental silos result in citizens being bounced around, without them in a place to know that they are trapped in a system that shouldn't operate as it does. Even when they realise, evidence of actions may be limited.

While front line delivery improvements constitute the primary stated goal of data sharing, the better use of data work consoles itself entirely with policy formulation, where real world difficult issues can get avoided towards irrelevance.

Registers must also be followed by citizen level transparency for services with a patient facing component - the citizen specific list of data uses, behind a system such as Gov.UK Verify, can cite the relevant entries in the Register. While useful for single releases, it is absolutely necessary for transparency of longitudinal studies, in a way which ensures high quality research can continue. While some studies have been effectively unavailable to the individuals within them in the past, transparency to each citizen about how their data has been used will allow even the oldest of studies a mechanism deliver on modern requirements (as applicable for that study). The history of the DWP Longitudinal Study shows how missing oversight is a problem, and the need for all research to be visible to data subjects. Ongoing publication, free to all, is the strongest defence on the integrity of the process in the face of the public - it demonstrates the benefits in a meaningful practical and substantive way, and closes a feedback loop from study to patient (which, in some cases, may have been non-existent for decades), and builds upon the benefits of modern open access to research publications.

The NHS now (mostly) understands the benefits of transparency - of telling data subjects how their records are used. The potential benefits from confidence and trust in medical research are very high; the barriers low. The same applies to the rest of Government.

medConfidential see no reason that NHS AI projects are inherently [controversial](#) in a world of consensual, safe, and transparent data use; they [could be routine](#).

2. Cite the law: Better decisions, and fewer scandals

In short, many data sharing questions come down to discrepancies in the answers to questions "*What powers are you using? What have you told your data subjects?*" Those are questions that every project must answer, and be seen to answer.

Like other IT projects, data projects fail for many reasons, and many deeply flawed projects are paid no attention whatsoever - they only become a scandal when the public are lied to. Care.data was paused because what the public were told was not true, and demonstrably contradictory. But it was cancelled because denial of those problems became infeasible.

"history also shows why computerization efforts ultimately ran counter to the modernization projects of the state, hurting industry and the nation at large".² However, a new approach to transparency is now possible because of widespread digital services - Gov.UK, Verify, and new legislation together give new solutions to old problems. Alongside registers of data sharing, that decisions are both minuted and findable leads to better decisions (in the long term, if not immediately).

Most mistakes around Whitehall are the result of good people doing what they think is the right thing with the best of intentions. Just as posting DVDs of data worked fine for years, until it ended very badly.

The constant drip drip of data losses and mistakes will steadily erode public trust in all data projects. If a citizen can see all the ways their data is used, the problems become seen in the context of all uses, not just problems. And with citizen focussed transparency, via Verify, when there is a problem, the citizens who are affected can be told what they need to know, and the vast majority of citizens who are (hopefully) not affected can also be told what they need to know. Knowing for certain that your information was not involved in a high profile incident is an act of reassurance - one that the media do not provide.

When all data projects and flows have some information published, a breach is one flow amongst many, and not an immediate systemic concern. While the arrival of AI is increasing the understanding of data amongst civil servants, it would be wrong to assume the same will not be true of the wider public.

Absent a clear grounding in evidence and fact, that public understanding may degrade to the sum of all fears, rather than a digitally literate society using data and services with trust and confidence. In a policy decision making context, that translates to fearful half measures vs full transparency both to a individual citizen and society. When it chose well, the UK has led the way in legitimate public reassurance, such as in the fields of embryology and communications; but has numerous other examples of failing badly.

² Programmed Inequality: by Marie Hicks. MIT Press. ISBN [0262035545](https://www.amazon.com/Programmed-Inequality-Marie-Hicks/dp/0262035545)

3. Citizen accountability: Increase the effectiveness of other reforms

“We have learnt from the tech sector that expertise needs to be at the heart of strategy. Relying solely on the well-meaning generalist, which has not served government policy well in computer science since the 1950s, is not enough.” - [Sir David Ormand](#)

Current data practice in Government is 5 different scandals away from adequacy. As shown by our work in the NHS, some of those will be public, some of those will be private - the scandal is optional, the improvements are inevitable.

It is easy for mediocre middle managers to again believe the sales pitch to buy a [big system](#) that will “do everything” - to flood a data lake. Such an approach has the benefit of looking like you’re doing something, while leaving the failures to be their successor’s problem. However, those bills have started to become due faster, and poor value for money and ineffectiveness also come with costs.

Bureaucratic inertia exacerbates the tension between the eternal desire for a copy of the data, and the institutional safeguard of APIs. Whether the fashion for “digital transformation” survives, or is superseded, the data controller and the data subject will endure, and there must be an ongoing communications mechanism between them.

Public services face dual pressures - to use data more, but also never to use data badly. In the current environment, these pressures are irresolvable. However, in a system of flow controls, there will be a feedback mechanism to account for where data was not used but should have been, or vice versa. Public registers of data flows will allow interested parties to engage on an ongoing basis with proposals and plans, while individual citizen level data flows allow a citizen to see how *their* data was used - or more commonly, how it was not accessed when it should have been.

There will always be countervailing voices, however, the replacement for the data ethics framework should facilitate an informed debate.

If there is an informed citizenry, with citizens able to see and understand how their data has been used by government, the more complex questions of AI and algorithms in the public sector become tractable. If an understanding of the status quo may lead to a collapse in public services, change is impossible. This implies that limits on accountability around data will increase the political pain that Ministers will suffer as they try to deliver an agenda.

When a citizen understands how their data is used today, it is far simpler to understand and explain the changes and improvements that will come tomorrow.

medConfidential

coordinator@medConfidential.org

October 2017