care.data: a Roadmap for Whitehall or a Warning Sign to Treasury?

Building an information economy, making the UK not just fit for the 21st Century but globally competitive means that data programmes and infrastructure will underpin and make key contributions to each one of the five spending round priorities.

Many of these programmes will, in some form, require the processing of bulk personal datasets[1] – be those NHS medical records, HMRC / DWP datasets, the National Pupil Database,[2] or population-scale transactional or administrative datasets used by other Departments.

Such programmes are almost always high profile or high impact, sometimes both. Poorly implemented, they can incur financial, even economic, cost – including significant economic opportunity loss – as well as reputational damage in Whitehall or Westminster, and consequent harm to public trust, public institutions and the nation at large.

The Prime Minister's Challenge Fund, for example, is a pot of money "to improve access to General Practice".[3] Funding in wave one was used to encourage digital interactions, or to digitally assist and underpin interactions between GPs and patients, e.g. via online consultations. In June 2015, realising rather late in the day the need for data to establish an evidence baseline, NHS England decided to use powers to take a copy of every GP's appointment book,[4] then walked this decision back,[5] and then further confused the situation to Parliament,[6] misrepresenting what it was doing both before and after the reversal.[7]

care.data is another "high priority" project of the NHS and the Department of Health,[8] intended to provide a detailed, linked medical and social care history of every patient and service user in England, solely for uses *other than an individual's' direct care*.[9] Such a database is highly sensitive, given it will contain health events of every man, woman and child in England, potentially going back to the start of computerised records in the NHS – but certainly, going forward, accumulating a new, centralised lifelong health record.

In 2012, a previous attempt was made to extract data from the GP IT suppliers' systems in a similar manner to the PMCF example above. When it surfaced in the summer of 2013, the care.data programme first put some posters and leaflets in GP surgeries, and then – having failed

[1] "Bulk Personal Datasets are large databases containing personal information about a wide range of people." - Intelligence and Security Committee of Parliament, 2015 'Privacy and Security: A modern and transparent legal framework'. http://isc.independent.gov.uk/files/20150312_ISC_P+S+Rpt(web).pdf

[2] https://theodi.org/consultation-responses/proposed-amendments-individual-pupil-information-prescribed-persons

[3] http://www.england.nhs.uk/ourwork/qual-clin-lead/calltoaction/pm-ext-access/

[4] http://dailymail.co.uk/news/article-3168803/Privacy-storm-GP-visits-No10-demands-details-millions-confidential-appointments.html

[5] http://www.dailymail.co.uk/news/article-3170204/NHS-backs-private-data-Demands-millions-confidential-files-abandoned.html

[6] http://www.england.nhs.uk/wp-content/uploads/2015/07/letter-dr-sarah-woolaston-mp.pdf

[7] https://medconfidential.org/wp-content/uploads/2015/08/2015-07-24DearMrCameron.pdf

[8] Q631, Tim Kelsey to Health Select Committee, 21 January 2015 http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/health-committee/handling-of-nhs-patient-data/oral/17740.html

[9] i.e. care.data does not directly affect the care you receive. It is solely for 'secondary uses'; purposes such as commissioning, "healthcare intelligence", research, commercial re-use, etc.

to meet even basic fair processing – spent a further £2m on a "junk mail leaflet"[10] campaign at the beginning of 2014.

The programme was "paused" shortly after NHS officials categorically denied having sold 25 years of medical events[11] to insurers, only to have some insurers thank them in a report that emerged a week or so later for doing precisely that. (The database in question is called Hospital Episode Statistics[12], or HES – though as an accumulation of dated, linked patient-level episodes, the contents of HES are definitely not statistics.)

DH / NHS had presumed consent for the Hospital Episode Statistics since 1989; when the public found out what was being done, the reaction was swift, predictable, and broad. 18 months later, DH is still working on resolving the issues.

While the NHS is one of the areas at the forefront of public discussion around bulk personal datasets, such cases are likely to be indicative across Government and in the public services in particular.

Data debacles such as care.data, and the ongoing and growing "data trust deficit"[13] are damaging to the UK economy – but evidence shows this need not necessarily happen. The Administrative Data Research Network[14] (BIS/ESRC), for example, is already constructing a network of managed remote access safe environments hosted within academia, enhancing a model proved by ONS. This approach to handling bulk personal (and business[15]) datasets needs to be scaled more widely.

A (Data) Trust Deficit within Whitehall

Unfortunately, when it comes to information, public bodies tend not to trust other public bodies. This being the case, bureaucracies will require every scrap of data they may ever possibly need in order to justify anything - though in many if not most cases, will never use much of it.

It would be entirely possible, for example, for the DWP to trust the NHS's assessment of a citizen's health status and, where the citizen consents, simply provide DWP with the particular 'statement of health status' that it needs. However, this is not how the Work Capability Assessments were designed.

No department trusts any other department, no organisation trusts any order organisation, no cubicle trusts any other cubicle. Attempts at transformation using mass data sharing to hide poor process and institutional inadequacy are unlikely to succeed.

---

[10] http://www.theyworkforyou.com/debates/?id=2014-02-25c.146.9#g148.5

[11] A "medical event", of which there have been about 1.1bn since 1989, as recorded in the "Hospital Episode Statistics". HES is not statistics as normally considered, but consists of rows of dated health events linked via an identifier for each individual. In ONS parlance, HES is highly disclosive linked microdata.

[12] http://www.hscic.gov.uk/hes

[13] www.statslife.org.uk/news/1672-new-rss-research-finds-data-trust-deficit-with-lessons-for-policymakers

[14] http://www.esrc.ac.uk/research/major-investments/Big-Data/BDN-phase1.aspx

[15] http://www.ons.gov.uk/ons/about-ons/business-transparency/freedom-of-information/what-can-i-request/virtual-microdata-laboratory--vml-/index.html

For integrated service development and, as a side-effect, greater public trust in the usage of data, far more use should be made of the publication of Official Statistics. Properly structured, designed with diversity to be resistant to 'gaming', properly-constructed statistics are far more flexible and safe than sharing bulk personal datasets. They also allow Departments to show the evidence base for certain decisions.

While the different bits of Facebook trust the data and processes of other bits of Facebook, the same is not true of Government.

Government Data Sharing isn't like Facebook Sharing

The relationship of the public to Government is very different to that of their relationship with Facebook or Tesco. Citizens themselves choose what information to post to Facebook or Twitter; they choose when to use their Tesco Clubcard. (And beyond this, they have a choice of supermarket or social network as well.)

While these services may receive a significant amount of data, there is no compulsion on their part and the choice is almost entirely on the side of the individual. An individual chooses what to post to Facebook or Twitter, and who can see it; a problem Facebook solved even before 2009.[16]

Government 'services' are monopolies, supported by the statutory enforcement powers of the State. To deliberately omit information from a submission to HMRC is a crime which can lead to imprisonment. Gmail may require a phone number to sign up, Facebook strongly encourages one, but no-one is required to have even a telephone number in order to deal with Government. Facebook might delete your account for having an odd name, but it cannot put you in jail. Facebook is for whoever it wants to be for – public services do not have that luxury; they must be for everyone.

A private company can do whatever it wishes within its terms of service, and users have only the choice to walk away. Government, and government 'data sharing', does not have the luxury of allowing only citizens who agree with its processes to use it. Any data programme that baldly claims "data sharing" as the sole solution to a set of complex interacting issues, or seemingly intractable problems, should probably be given extra scrutiny – especially as regards what data-minimising alternatives there might be.

Regaining and Maintaining Trust in How Data is Used in Government

The Royal Statistical Society has described the "data trust deficit" – which is that, regardless of sector or organisation, there is a measurable difference between the perception of trust in an organisation generally and trust in that organisation with regard to its handling of personal data. This problem is widespread; it is fundamental, and it is systemic. It can also be solved.

A previous Government standard for data transfer used to be posting unencrypted CDs; then the HMRC lost some discs[17], and the standard changed. It will change again, to maintaining all

[16] http://news.bbc.co.uk/1/hi/uk/8134807.stm

[17] https://en.wikipedia.org/wiki/Loss_of_United_Kingdom_child_benefit_data_(2007)

population-scale bulk personal datasets within “safe settings”, and using remote access tools for users to connect to and process the relevant data *in situ* – removing only non-disclosive results, rather than passing copies of the data around. These mechanisms, safe settings, also allow for reporting back to citizens who’s had access and how their data has been used.

In 2012 the Department for Education attempted to make the bulk personal dataset containing the linked school history of every child in the country available to all. That scheme proved unwise and was dropped, but it did raise awareness amongst privacy-sensitive constituencies, who were concerned about the purposes for which the data had already been used.

A copy of that list of purposes was requested under the Freedom of Information Act and, when published, many otherwise concerned parents were reassured. Unless there is active wrongdoing or misuse, the more knowledge[18] individuals have about what is happening to their own data, the more reassured they tend to be. As with breach reporting, people’s fears and concerns are highly individualistic – and can only be contained by knowledge of whether or not you were affected and, if you were, what is being done about it.

Giving citizens a complete list of how their data is used is the only measure that will build confidence into the long term, and resilience. A Government Data Usage Report[19] in the form of an on-demand digital report – available only to the citizen to whom it pertains, using mechanisms made possible only because of the move to “digital by default” services – can provide citizens with usable evidence that Government is trustworthy.[20]

“If you don’t agree to share data, you may not receive…”

If a citizen chooses to not to share data, there should not be any proactive denial of anything; but Government may legitimately point out that if an individual doesn’t agree to some data sharing, they may have to repeat themselves to other bodies.

In practice, when having to repeat yourself because you didn’t share information is the main outcome, this proves to citizens that their information wasn’t shared against their wishes. It may make a digital transaction have a couple of extra steps – as more information may need typing into a form – but it does not meaningfully affect service delivery. Some will choose convenience, some will choose otherwise; each for their own reasons.

Meanwhile some current and ongoing approaches, the “phrased like a threat”[21] legacy of the data programmes of the last Labour government, reveal vestiges of the thinking that supported the ID cards programme.

[18] https://www.whatdotheyknow.com/request/requests_for_access_to_national#incoming-340035
[19] e.g. https://medconfidential.org/wp-content/uploads/2015/08/gov-data-usage-report-april-2015.pdf
[20] See Baroness Onora O’Neill on trust vs trustworthiness: https://www.ted.com/talks/onora_o_neill_what_we_don_t_understand_about_trust/transcript?language=en
[21] Health Select Committee, Oral evidence: Handling of NHS Patient Data, HC 1105, Tuesday 8 April 2014: http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/health-committee/handling-of-nhs-patient-data/oral/8416.html

The benefits to the UK

The UK has the benefit of European-style Data Protection laws within an ethical governance framework, and a US-style innovation culture. As such, given clear guidance, the UK can lead the world by honouring both aspects and making data and information systems work in the interests of all.

The NHS has already announced that it is working on a mechanism for a patient to know everywhere their data has gone.[22] It is those data debacles at the opening of this paper which have necessitated this solution, in order to restore patient trust – which many will ultimately be happy with, because the vast majority of data flows within and across the NHS and wider care system are entirely within their interest. But in the current information and trust environment, citizens only hear bad news; they (and Government) have no way of finding out whether any one piece of bad news affected them and, if it did, what they can do about it.

The Government Digital Service (GDS), and features of GOV.UK Verify, have the potential to fundamentally reform Government processes and trust in a manner which "more data sharing" does not. In the NHS, the reformed Health and Social Care Information Centre is developing into the role of trusted custodian, providing reporting to patients; in the rest of Government, this role will need to be facilitated by ONS (delivered via GOV.UK).

As the current spending round decisions get made, what gets measured will get managed. Whenever data programmes and projects are proposed, there should be serious scrutiny framed by the questions raised above and summarised below, to ensure that the costly NHS experience proves to be a cautionary tale, not a handbook.

Properly done, there is the dramatic potential for data programmes and infrastructure to lead to more effective and efficient public services; improperly done, it will cost dearly.

Questions to ask of *all* Data Programmes and Initiatives

1) What are the unintended consequences of your project?
2) How are you seeking (and refreshing) citizens' consent to do this?
3) How will citizens know what has happened with their data?
4) What is the project's public trust and economic opportunity loss potential?
5) Which decisions of other parts of Government have you chosen to ignore?
6) How will this project contribute to increased trust in the data policies of HMG?
7) What are the process transformations that go with data transformations?
8) How will this project start to close the Data Trust Deficit? Or will it make it worse?

Answers are unlikely to be comprehensive at project initiation, but they should at least be considered and published alongside the proposals.

[22] "By June 2015, the HSCIC will develop proposals with industry for personal data usage reporting." See timeline, p58, *Personalised health and care 2020: a framework for action*: https://www.gov.uk/government/publications/personalised-health-and-care-2020