Written evidence submitted by medConfidential (BIG0088)

1. medConfidential is an independent non-partisan organisation campaigning for confidentiality and consent in health and social care, which seeks to ensure that every flow of data into, across and out of the NHS and care system is *consensual, safe and transparent*.

2. Health data was, and is, the canary in the big data coalmine of public consciousness, both in the public sector and beyond. When it started chirping, NHS England dismissed it as just tweeting, ignored it, and blundered on anyway. The ignominy of care.data is the result of that approach. It was avoidable, the Senior Responsible Owner of Care.Data at NHS England, Tim Kelsey[1] actively chose not to avoid it, preferring to maintain the underlying culture of the project even until now; described as "what is the most data that we can sell, without the public objecting?"

3. The objections of the public, and the Institutional response has been bifurcated, with the two public bodies taking radically different approaches. NHS England blundering on regardless, HSCIC accepting public concerns, going through a process of introspection, and considering what a data centric public body should do in the 21st Century.[2]

4. The history of care.data is as follows. In the late 1980s, the NHS set up what is now the "Hospital Episode Statistics" containing data about all "events" in hospitals, linking patients over time. In the late 1990s, Tim Kelsey, then an FT journalist started a company to commercially exploit the data, entering into a joint venture with DH in 2005.[3] In 2011/2012, he joined NHS England to expand that approach to all data linked across the NHS. Firstly, they went to the IT companies that supply your GP with their patient records system,[4] and asked for a copy of all records; the companies "politely" suggested they might want to ask the GPs first. So then they asked the GPs, sending posters to go up in patient surgeries in August 2013.[5] That didn't meet the legal requirements, which the ICO said should be a letter to each individual patient. In January 2014, NHS England sent a "junk mail leaflet" to each household in England that had not opted out of junk mail.,, which contained no details on how to opt out, or much detail at all. Patients were given 6 weeks (or less) to opt out, and with 2 weeks to go, the Secretary of State pulled the plug. In those 4 weeks, between 900,000 and 1.4 million people had gone to the Internet, searched and found an opt out form, downloaded it, printed out multiple copies, hand filled them in, and walked them into their. GP surgeries. This was not a low barrier to action, an action likely taken by between 0.5 and 14.5% of the constituents of each member of this committee representing a constituency in England.

5. The only permanent solution to a systematic data dilemma is the ability for you to know what data an organisation has on you, know how they used it, and know where it went.

6. **Prologue**

7. midata was a project to let citizens access the data held about them. In practice, it was taking the paper based SAR process from the 1990s, and digitising it, but not the law, incentives or culture, for the modern world. It was done by consent not mandate, and didn't entirely work that well, as there was never the impetus or ecosystem around the process, although some tools were built which saved those who used them some money.

---

[1] Tim Kelsey announced his resignation from NHS England in September 2015, and leaves for Australia in December 2015.

[2] http://hscic.gov.uk/datareview

[3] http://www.publications.parliament.uk/pa/cm200607/cmselect/cmpubacc/368/368.pdf

[4] GP IT systems are provided by one of only 4 suppliers (EMIS, TPP, INPS, Microtest)

[5] https://twitter.com/medConfidential/status/431454828615577600

8. As an aside, this is a very different model to the " "Building trust, removing the friction"[6] approach of the so called digital catapult, which could be described as trusting those who buy it not to abuse it. A strategy that not only shows a complete lack of understanding of current data brokerage practices, but assumes that data predators will not pick on the most vulnerable. The catapult's position seems to be that as long as they get their cut, the social harms can be resolved by others. If any evidence to this committee should begin to sound the death knell of the parasites that predate on your vulnerable constituents, it may be that from the publicly funded digital catapult.

9. External agencies (either intelligence or advertising, depending on your poison) wish to be able to see anything that exists and mine it for their objectives, and will continue to do so until public concern generates restrictions. Whether that was the care.data debacle, or the sale of data on children in state schools to journalists,[7] the current data dilemma is "distribute as much as we can, and hope it doesn't go wrong." We commend the RNLI who recently made the decision not to sell their membership list.[8] It will be interesting to see the medium and long term response to that decision.

10. Organisations, and individuals, keep getting hacked - "talk talk" which affected 150,000 customers,[9] or more recently, "Vtech"[10] which leaked 6.5 million children's details, including photos and text message conversations. These losses of bulk personal datasets show that the primary cost of storing such datasets is not financial, or technical, but it is legal and reputational. New measures will be required.

11. The common failing is an organisation disregarding the fact that it deals with human beings. Whether it's the Department for Education forgetting about children,[11] NHS England forgetting about patients in care.data, or researchers forgetting that "Whitehall II" isn't just a database for research, it has living, breathing civil servants as the data subjects.[12] Every personal dataset contains human beings, and no matter how projects are spun, people are generally good at telling the difference between something being done "to" them rather than "for" them**.**

12. **Lessons from care.data**

13. Care.Data, and the furore around the Hospital Episode Statistics, was a problem because, fundamentally, it surprised the patient. There was a widespread expectation that confidentiality would be maintained by the NHS and the doctors that patients met and trusted, and people were horrified to find out it wasn't the case. Dr Ben Goldacre has phrased what he feels should happen with his characteristic clarity:[13]

> " 1. Keep patients' secrets.
> 2. Don't keep secrets from patients "

---

[6] http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/science-and- technology-committee/big-data-dilemma/written/24296.pdf
[7] https://www.whatdotheyknow.com/request/293030/response/723407/attach/5/The%20Times.pdf
[8] http://rnli.org/NewsCentre/Pages/RNLI-is-first-major-charity-to-give-supporters-control-over-contact1.aspx
[9] http://www.bbc.co.uk/news/business-34743185
[10] http://www.bbc.co.uk/news/technology-34980655
[11] https://theodi.org/blog/open-data-and-national-pupil-database
[12] http://www.bmj.com/content/351/bmj.h5817 reply to http://www.bmj.com/content/351/bmj.h5087
[13] https://twitter.com/bengoldacre/status/656577067698356224

14. Surprising the public, unpleasantly, when they have a choice, is likely to lead to that choice being made in a way an organisation would prefer not to be made.

15. Baroness O'Neill has spoken at length at the difference between "trust" and "trustworthiness".[14] Care.Data was a classic example of a bureaucracy confusing the two. Unfortunately, NHS England has not learnt the lesson, and Public Health England believes the lessons did not apply to it.[15]

16. Care.data was about creating a linked, lifetime, quantitative analysis of every citizen's life, covering the entire population. This allows for targeted attacks, where one unfortunate event in a person's life, which is reported[16] or otherwise identifiable, such as the date being public and it being a semi-rare event, being a key to all the other details held in their medical record.

17. The date of birth of children is not usually secret (it's institutionally sensitive, but their friends probably know it). The NHS rightly considers that the date of birth of a patient is sensitive, but does not consider that the dates of particular events are sensitive, where those events can include maternity events. As such, it is greater than 90% likely that any woman with two children can be identified from her individual level linked historical medical record (the sort the NHS has been selling for years). Where the woman has 3 children (ie, 3 different dates), it's pretty much certain. The emergence of big data is an important scenario there. When a Government department looked at some passport data, the "data science" specialist found a couple of individuals who arguably shouldn't have had duplicate passports[17], and a long list of individuals in Government employ for whom being so easily detectable raised national security concerns.

18. Badly anonymised data is not anonymised data, it is identifiable data. And being identified is somewhat like being pregnant: it is a binary state. And when you have gone from one state to the other, getting back can be a non-trivial operation, effectively impossible without widespread co-operation.

19. **Principles of a Solution**

20. The only long term, scalable solution requires full accountability to an individual of how their individual level data has been used. Parliament may decide that, on the balance, care.data 2 is a programme that should go forwards.[18]

21. Accountability of data usage and data flows will lead to better use. More transparency over where data goes, and better use of data, must go hand in hand. The institutional fear of the civil service is not that it will be revealed how they are using data, it is the reveal of how *little* they are using data, how decisions are actually made, rather than how it is portrayed that they are made. If those on Job Seeker's Allowance could see every time someone from DWP had looked at their

---

[14] https://www.ted.com/talks/onora_o_neill_what_we_don_t_understand_about_trust?language=en

[15] evidence forthcoming: http://www.parliament.uk/business/committees/committees-a-z/ commons-select/health-committee/inquiries/parliament-2015/public-health-post-2013-inquiry-15-16/

[16] http://www.telegraph.co.uk/news/election-2010/7633714/General-Election-2010-Nick-Cleggs-wife-fractures-elbow-in-shopping-fall.html

[17] data quality everywhere is terrible. The number of passports issued with a *country* of birth as "Yorkshire" would (possibly) astound you.

[18] it's barely better than their last attempt at this point, but is at least more transparent about the long term plans (substantively identical to those from last time), and there will be a consultation "early next year."

record, there would be a great deal more honesty in the system. What would the discussion between a claimant and their advisor be if the claimant knew who had looked at their record, when, and why? The new universal credit interface could easily have that as a tab on the screen for the individual to see - but it currently doesn't.[19]

22. This problem will continue systematically, repeatedly, and get worse, until every patient (in the NHS), or citizen (for the rest of Government), receives a complete account of how their individual level data has been used. In evidence to the Committee, the Director of Data at the Cabinet Office outlined the case for more data being used.

23. Does this committee trust that no decision that a Government bureaucracy makes will prove as unwise as NHS England's plans for care.data? Over the long term, that seems like a fundamentally losing bet to make, even if we take full regard of Baroness O'Neill's comments referenced earlier.[20]

24. The Confidentiality Advisory Group[21] at the Health Research Agency is concerned that, to do what they are charged to do, they are required to be omnipotent and perfect, based solely on the information that they are provided with. That would appear to be unwise.

25. The current model requires that trust is permanently upheld, and the Committee heard about the Data Trust Deficit from the Royal Statistical Society. Full reporting to everyone of how organisations use their individual level data, would mean that, as with your bank statement, where individual level data went was entirely accounted for. We attach a copy, for health records, of a Personalised Data Usage Report, which is what this could look like. As a result of the care.data fiasco, the NHS has begun to keep track of where it sends data (in that, the HSCIC has a capability to track down to an individual level, but other areas of the health system, such as Public Health England, still fail to recognise the public concern around care.data).

26. **Draft Investigatory Powers Bill**

27. Clause 150(3) of the draft Investigatory Powers Bill, seeming superficially for entirely practical reasons, treats the data of people who are deceased identically to those who are alive - the Data Protection Act allows them to be treated differently. This is a lesson for all considering big data, especially those who interact with the public, and the effects on families. While this may have consequences adversely affecting transparency around the Agencies, there may be value in deeper consideration.

28. **New ethical challenges of Big Data?**

29. If there is a new ethical challenge, it is whether or not, in a world where it is possible to tell a citizen everything that happens to their individual level data and why, we choose to do so. There can be only one ethical answer to that question.

December 2015

---

[19] https://twitter.com/smithsam/status/672786114688733185

[20] https://www.ted.com/talks/onora_o_neill_what_we_don_t_understand_about_trust?language=en

[21] http://www.hra.nhs.uk/resources/confidentiality-advisory-group/