coordinator@medconfidential.org
19 February 2025

Dear Professor Sir Ian,

We hope ONS teams will continue to deliver the public confidence you described when speaking at the Institute for Government in 2021 where you said:

> "There's no god given right for us to have data. There needs to be a really sound public good reason for collecting data, and using data, and people need to feel absolutely comfortable that their data are being used properly and kept securely and in a way that satisfies all forms of privacy."

It is unclear whether you or ONS still believe that. You recognised to the Treasury Select Committee that you can ask people to permit linkage of their data, but the approach of ONS seems to be prelink and not ask people at all.

Special pleading and cognitive dissonance are nothing new for the ONS (or NHS) to hear, but it was somewhat novel for ONS to be doing that itself. Across many conversations about data linkage and the IDS at the recent Assembly, the advocacy was often "there should be strong rules *but only for everyone else*". The suggestion that public confidence mattered to linkage is welcome, but the claim was that it could be disregarded for the linkages advocates wanted themselves, as the recent HDR/Sudlow Review also advocated for itself.

We welcome ONS publishing the list of projects in the IDS in advance of the Statistics Assembly (and the Assembly itself was a great success), and doing so allows informed debate only to the extent that those lists are accurate, descriptive, and timely. The current lack of clarity on which datasets are being used by a project suggests ONS is still on a journey to demonstrating trustworthiness and quality.

Remits of SRS and IDS

Today the IDS has a dual mandate of supposedly enforcing rules while simultaneously begging departments for more data and is entirely open to an official walking across "the Home Office building" demanding a project gets data – there aren't even any doors to slow them down.

The SRS had one job, and it should continue to have only one job: to be the trusted research environment which enforces stated rules on data access. SRS can continue to receive data files for safe projects, operate the environment, and run support services like output checking. The IDS should not have responsibility for the environment but solely be a service which allows *people* to be linked across datasets, respecting objections where applicable, but not managing the environment itself. When a project wishes to use linked data, the (cached) source data approved by data controllers and the linkage file are provided to the SDS for use, probably with an (IDS owned) standard ingest process which does the common data linkage into a single file for researcher efficiency. The IDS then remains responsible for linkage, the SRS remains responsible for ensuring the rules are followed, and the two are not in conflict. Merging both into one service is embedding conflict of interest by design.

ONS are Trumping clubcard culture

That the public increasingly do not wish to respond to surveys does not mean ONS can just take data on 100% of a relevant population while acting like those people gave that data willingly, when in fact no one has any choice, involvement, or even awareness at all.

It is a fundamental mistake for ONS to treat admin data as akin to survey data, and so the ONS position is incoherent. The collection mechanisms and powers are radically different, and when ONS argues that ONS is trusted with survey responses so it should be trusted with admin data, it ignores the widespread concern that the LFS, the flagship survey, now has a response rate of 20% even when making promises to respondents which are entirely lacking when just copying admin data. Talking to your team around the Assembly and elsewhere, the question I repeatedly asked was where is the equivalent to the NHS's National Data Opt Out? When legitimate controversy starts on admin data, what does an outraged citizen *do*?

The current position appears to be that a citizen can do nothing and should accept both the National Statistician and Government more widely doing anything it deems appropriate solely because they deem it appropriate. We'll address the UK Biobank's implementation of that principle later, but the more settled precedent is the cancer registry who believed nothing could go wrong with a "causes of cancer study", only for the study to be from a tobacco company. Wholescale reform was required after that particular failure of governance. Without change, ONS will instinctively defend a practice that causes legitimate anger, and ONS's reputation across all of your remit will measure the consequences. For now, you can look at the response rate to the LFS and similar to see what the public may do.

How might ONS explain just taking data to the public?

As a thought experiment, we envisaged to your predecessor what the National Statistician would say on a card that an interviewer left with a survey respondent thanking them for volunteering to do their part to support the work of the ONS. We enclose an updated example (the green one).

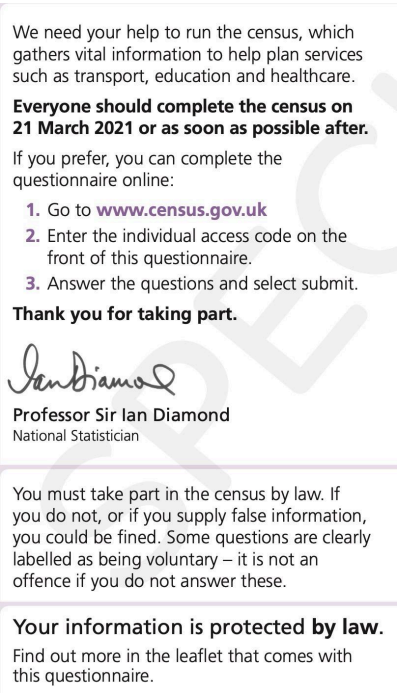
What would the equivalent be for admin data – what would you want to put on a card you handed to a data subject about the admin data you take today, or under the proposals on page 123 of the HDR/Sudlow Review? We include our example (the red one).

You may wish to ask your team what they would write on their version, and whether it reflects current practices, and what the public would say if (or when) it did reflect current practices.

The working assumption of your colleagues seems to be that because ONS takes any admin data it wishes for any purpose it sees fit, it need not consider the impact on ONS reputation and the consequent response rates for surveys.[1] As you argue so well, there is a lot of information you can only get from surveys, and that requires the public's ongoing cooperation.

[1] Relatedly, it is the position of DHSC (and the top left of page 167 of the HDR/Sudlow Review) that the promises made to respondents during the pandemic about uses of GP data should be simply torn up and the new Secretary of State should do what the then Secretary of State said he wouldn't. It is unclear what the effect of that change will be on various ONS efforts.

Census data is protected, admin data is not


We need your help to run the census, which gathers vital information to help plan services such as transport, education and healthcare.

Everyone should complete the census on 21 March 2021 or as soon as possible after.

If you prefer, you can complete the questionnaire online:

1. Go to **www.census.gov.uk**
2. Enter the individual access code on the front of this questionnaire.
3. Answer the questions and select submit.

Thank you for taking part.

Professor Sir Ian Diamond
National Statistician

You must take part in the census by law. If you do not, or if you supply false information, you could be fined. Some questions are clearly labelled as being voluntary – it is not an offence if you do not answer these.

Your information is protected by law.
Find out more in the leaflet that comes with this questionnaire.

It is clear from the ONS practices and data handling around the census that ONS recognises that a 100% dataset from mandatory collection is something that is highly sensitive. The front of every census form has a request and assurance from you, as National Statistician, about how the data will be used and protected.

When a person fills in a survey or the census, you (or your representative) give some assurances how the data will and will not be used. ONS asks people what crimes they have committed and experienced – potentially the worst events in their lives – and people answer. Part of the reason for this is they are in control of their answers. Admin data takes that control away.

With admin data, there is no assurance, and it appears to be the belief of ONS that it can use any data it can get any way it wishes if it satisfies your definition of "public good". If ONS were to treat the 100% census data the way it treats 100% admin data there would (rightly) be mutiny in Titchfield.

The ONS target for census response was 94% overall and it achieved "at least 88% in every LA", yet 95% coverage of NHS health records when respecting the NHS National Data Opt Out is apparently a calamity in every case. When using NHS data for research and planning ONS can respect the National Data Opt Out and still be using better quality data than it gets from any other source.

In practice, ONS seems to believe it can take everything on everyone held anywhere without any other disclosures or individual choice – if someone reports being a victim of crime the HDR/Sudlow Review argues ONS should link that report to everything else in their life for arbitrary purposes, possibly by the perpetrators of the crime.

ONS's working assumption appears to be that the public discovering "even if you had a NHS Data Opt Out, ONS linked your health records to your tax records using your name/address/DoB and the results were used to (de)prioritise hospital operating theatres"[2] will be entirely unconcerning to the entirety of the public, and the primary methods of doing this are by ignoring the choices they do have and hoping no one notices. medConfidential can not fully assess the feasibility of ONS's working assumption that the public will remain in the dark indefinitely, but ONS's position may become untenable when scrutiny comes; and scrutiny always comes.

The HDR/Sudlow Review launch included a call for applicants, including eugenicists, to receive data in "days" proclaiming Biobank have "one of the best systems" for giving data out rapidly. Events and independent scrutiny show that to be folly as it had emerged (weeks before the launch) that biobank has "safeguards" so weak they let an organisation run out of the same dodgy registrar as QAnon front companies[3], supposedly because the applicant wasn't on the list they say they checked. It remains to be seen what happens to those who are on a list. Biobank continues to insist that their users are all "safe people" doing a "safe project" to have access within what they describe as their five safes environments. Would ONS agree with such an assertion? An external commitment closer to the

[2] https://www.england.nhs.uk/2024/12/world-leading-nhs-trial-to-boost-health-and-support-people-in-work/
[3] https://www.wired.com/story/epik-domain-registrar-new-owner/

census than the back room copying of admin data into the IDS means addressing the clubcard culture of government and dispelling the cognitive dissonance that permeates the shared culture of HDR and Biobank

The word games undermining the five safes

The "Five Safes" were designed by ONS and Felix in a largely benign world. However, any measure that becomes a target ceases to be a good measure, and the "five safes" ideals are now getting gamed. HDR pours budget into undermining them – whether that's their "federated TRE" which has a backdoor/"feature" for commercial users to run analyses that the TRE admin can not examine,[4] or allowing anyone with a linkedin account to run arbitrary queries on individual data.[5]

HDR has previously wanted to create a "sixth safe" to undermine the previous five, for various contrived reasons.[6] HDR's justification was the researcher "dream" that they perform an analysis so important that all safeguards must be skipped, so important that patients must change their treatment right away, but so unimportant that there need be no publication of results (even pre-prints), no replication, no publication of code (even for validation), nothing must get between a PhD student with an analysis earnestly insisting someone else's patient must change their treatment in a panic. A fictional scenario that is entirely nonsense, like projects to dilute the five safes model through "standard language" which will allow words which mean one thing in the context of ONS to be used to mean far less to others. ONS may never envisage that scenario, but it is what some want to use the IDS is for (and across the Atlantic, what the new US Government may use it for).

Your "colleagues" playing word games to mislead others and the public will harm everything ONS stands for. For example, on page 163 of the HDR/Sudlow Review, those that "would be well positioned to lead on SDE standards" are only "HDR UK [legal entity: HDR], ADR UK (partner with HDR but legally part of UKRI who fund HDR), the UK Health Data Research Alliance [legal entity: HDR], and "UKRI's Data and Analytics Research Environments UK (DARE UK) programme" [legal entity: HDR]. "PEDRI" and "PRUK" are other brands within the HDR hierarchy which promote the HDR corporate view without disclosing they are part of that hierarchy. The HDR/Sudlow Review suggested only listening to HDR.

Institutionally, HDR and Biobank share a culture which is to only listen to those who agree with them.[7] Like ONS, HDR has sought to make decisions itself about how NHS patient data is used reflecting only their institution interests rather than patient interests. We hope ONS is not degrading the same way.

---

[4] A "Deception by Design" project previously called TREEHOOSE, and since renamed several times.
[5] Previously called COCONNECT and HDR continue to lobby for their blackbox to be added into various TREs including ONS despite: https://www.youtube.com/live/TETh7aUfky4?feature=shared&t=1514
[6] See page 62 of version 2 of the D1 report https://dareuk.org.uk/how-we-work/previous-activities/dare-uk-phase-1-sprint-exemplar-projects/priam-privacy-risk-assessment-methodology/
[7] HDRUK's role as a private charity is to lobby, and they're very well connected to do it – we note here that it is HDRUK's corporate policy to refuse to engage with medConfidential (even if many of their staff and advisors recognise the folly of their CEO's insistence). You may wish to ask for a copy of the "dossier" which Andrew will happily share arguing how unreasonable medConfidential are for asking some (unanswered) questions about HDR's publicly funded activities: https://medconfidential.org/wp-content/uploads/2024/08/HDR.pdf

Available next steps: treat unsampled admin records like unsampled census records

ONS has to get to where it wishes to be from where it is today. It should be ONS policy that raw admin data is treated as if it were raw census data, and any divergences in treatment of one from the other must be transparent and justifiable. There are some differences, but a file of 100% admin data is much more akin to 100% census data than it is different.

Firstly, the ONS should respect the National Data Opt Out for health data projects which do not need 100% data – if a project would not equivalently qualify for the 100% census data, it should not qualify for the 100% NHS data (as with census, there will be some narrow projects that do qualify). The Department of Health in England may tell ONS that you do not *have* to apply the NHS National Data Opt Out according to their policy, but their policy is a minimum, not a maximum. If ONS can be forced by departments to process data it does not wish to process, that opens a much greater range of concerns. As with census data for today, if a project has done an analysis on the 20% dataset and can statistically demonstrate and explain why they need more granularity for a particular analysis, there can be a process to approve exactly that analysis (the NHS already has such a process).

Secondly, ONS should consider which part of Government should lead on commencing discussions about a "rest of government data opt out" (using the Gov.UK One Login to allow people to tick a box as the NHS NDOO does inside the NHS). You said to the Treasury Select Committee that sampling for surveys is becoming far more difficult. As DSIT rolls out the Gov.UK Account and Login, requiring both a currently valid phone and email, and potentially the provision of identity information, Login allows a new form of opt in sample for ONS to use and match to other sampling frames. Alternatively, the mixture of Login requiring phone/email and identity, the population index, and secretive population scale data matching may turn the entire thing toxic. This is where, a month ago, we would have suggested engagement with the Privacy and Consumer Advisory Group (aka OLIPAG) of GDS; however DSIT have abolished it. Someone in government will get greedy in pursuit of their own public task, and any claims ONS wishes to make about public interest efforts will get tested.

With your experience of the civil service, you know how that will go, and organisations are already started playing the word games that will inevitably undermine the five safes model.[8]

Yours sincerely,

Phil Booth, medConfidential Sam Smith, medConfidential

[8] One of the reasons this letter has taken so long to send is waiting for various items to become public. Sod's laws means they may appear before you read this, although your friends may have briefed you on what they've done…