Analysis of the NHSX Contact Tracing App 'Isle of Wight' Data Protection Impact Assessment

Dr Michael Veale, *Faculty of Laws, University College London*
9 May 2020

Summary

This note examines the published data protection impact assessment (DPIA) released by NHSX in relation to their contact tracing/proximity tracing app.[1] It highlights a range of significant issues which leave the app falling short of data protection legislation. It does this in order so that these issues can be remedied before the next DPIA is published. The main issues this note focuses on are the following:

Personal data

- **The DPIA must not claim this data is anonymous, or that the app preserves anonymity, as under UK law, it does not.**
- **The document (and associated public messaging) must be changed throughout to reflect the fact that it is not the case that personal data about a user is only uploaded with a user's permission, as *other people* upload data revealing a user's social interactions.**

User rights

- **The lawful basis for a blanket refusal of the right to erasure is unspecified by NHSX in this DPIA.**
- **The NHSX App unlawfully designs out the right to access when there is a legal obligation to design it in.**
- **If the controller plans to, as with the right to erasure and the right to access, refuse all attempts at the right to object, this needs a justification in the DPIA.**

Monitoring and automated decision making

- **The DPIA must acknowledge the NHSX App systematically monitors publicly accessible spaces**.
- **The DPIA does not set out a valid lawful basis for the solely automated, significant decision-making it correctly identifies as occurring.**
- **The information contained in the document embedded in the DPIA describing the logic of automated decisions must be provided under GDPR, article 13.**

Prior consultation and e-Privacy

[1] Data Protection Impact Assessment NHS COVID-19 App PILOT LIVE RELEASE Isle of Wight (6 May 2020). Retrieved 8 May 2020.

- **The Information Commissioner must be consulted prior to processing within the meaning of GDPR, art 36, not just briefed.**
- **The DPIA should explain how the The Privacy and Electronic Communications Regulations are complied with, both in relation to Bluetooth usage and in relation to embedded trackers.**

The note does not consider alternative architectures or less intrusive means to achieve the purposes of the NHSX app, although these are critical issues that this DPIA could be argued as failing to assess.[2] This note is unable to assess the risks of the app as provided by the DPIA as all the risks have been redacted.

Contents:

Data is not anonymous

Collection of Personal Data is not always a Voluntary Action

Systematic Monitoring

Users deprived of data protection rights

- Deprivation of the right of erasure (GDPR, art 17)
- Unlawful deprivation of the right of access (GDPR, art 15)
- Right to object
- Deprivation of rights requires a legislative provision

Automated decision-making

- Processing under data protection legislation does not include decision-making
- Processing under the Health Service (Control of Patient Information) Regulations 2002 does not include decision-making
- Logic of automated decisions

Risks

- Article 36 Prior Notification

The Privacy and Electronic Communications Regulations

Data is not anonymous

The DPIA is a document relating to data protection law, yet consistently misuses the terms ‘anonymous’ and ‘anonymity’.

In several parts of the DPIA, NHSX state that the system preserves anonymity:

[2] See generally Matthew Ryder QC, Edward Craven, Gayatri Sarathy and Ravi Naik, *COVID-19 & Tech responses: Legal opinion* (Matrix Chambers, 3 May 2020).

> The App is designed to preserve the anonymity of those who use it. It does **not** collect any directly identifiable information (for example, it does not collect name, telephone number, NHS number or GPS location data). (p.3)

> This will not involve the disclosure of information that reveals users' identities back to the App, and this process will continue to preserve users' anonymity (p.3)

> The App is designed to preserve the anonymity of its users. (p.6)

> This information will be used anonymously to encourage anyone who has recently come into contact with you, and has this app installed on their phone, to self-isolate. (p.8)

These statements are legally misleading, and contradictory to later admissions in the DPIA. The NHSX app does not preserve the anonymity of users, as it primarily processes pseudonymous, not anonymous, personal data. Anonymous information is only that which is not personal data.[3]

The DPIA later concedes this, and makes contradictory statements:

> The processing of data including identifiers that are unique to individuals (which therefore meets the definition of personal data) (p.7)

> Data may be fully anonymised for public health planning, research or statistical purposes, if those purposes can be achieved without the use of personal data. (p.10)

> The data collected by the App is pseudonymised – albeit having never been directly identifiable there is no lookup to users' identities. (p.11)

> Through Sonar ID and other related identifiers, the data allows the individuation of users. However, it is processed in a form that in cannot [sic] be attributed to (i.e. reveal the identity of) a specific data subject without the use of additional information, so we are treating the data as pseudonymised.

> Will the processing involve a large amount of personal data (including pseudonymised personal data) and affect a large number of data subjects?: Yes (p.12)

> At no point is a user's identity capable of being captured or disclosed by the instance of the App installed on a phone, or the central database. (p.3)

> For this reason, we are treating the data as pseudonymised data – albeit it has never been capable of revealing an individual's identity. On this basis,

[3] GDPR, recital 26.

> although an individual will not be identifiable from the data, the data will qualfy [sic] as personal data, and the GDPR applies. (p.3–4)

The data in the NHSX app *is* 'capable' of revealing an individual's identity. Whether NHSX *intend* to do this is not a relevant question from a legal standpoint, the question is whether it reasonably could. This can be (non-exhaustively) illustrated by three several simple, plausible scenarios. Other more advanced technical methods for reidentfying social graph data through analysing its structure are possible, but not necessary to discuss here.

1. An organisation collaborating with the NHSX server places a mobile phone/Bluetooth receiver they control at a point where an individual's identity is declared, such as a passport booth, or an Oyster card reader. Both properties and surrounding infrastructure are owned by the public sector, and may even contain sensors that can be easily repurposed. By a simple process of matching, the identity as disclosed by the passport/travelcard/similar can be matched to the persistent identifier in the system.

2. Sonar IDs are stored next to part of their postcode. Take a travelling individual A. Other individuals that they pass in the street, or potentially even while driving past houses, will both emit their identifiers and record A's identifier. A travels to several places (either over the same day or multiple days), and post information of his journey to a social network. Particularly if these places are not adjacent (e.g. if A travels far), it is likely that A's journey is unique for those days — that no-one else in the UK went to Taunton on Tuesday, then Reading on Wednesday, then Cardiff on Friday. A will encounter a disproportionate number of people with the local postcode in each of these areas. If A uploads their information after testing/declaring positive, A can be uniquely discovered in the dataset from those three postcodes. If A encountered people who later uploaded *their* information after testing/declaring positive in each of those three places, A's identity can be discovered without A having ever consented to uploading their data.[4] As an additional confirmation, A's identity may even be able to be cross-referenced with the easily-accessible EXIF data in photos they have uploaded to social networks, which can confirm the make and model of their phone — variables also stored on the central server.

3. The Sonar ID on the device could be extracted by police from the app through a 'cyber kiosk'.[5]

- **The DPIA must not claim this data is anonymous, or that the app preserves anonymity, as under UK law, it does not.**

[4] The DPIA states 'The App will not collect data about which postal district a user might be in from time to time as they move around' (p.9). However, in this case, there is no need for the App itself to collect such data. The inference is trivially made on the backend server, because all Sonar IDs are associated to a postcode-half, and Sonar IDs are associated to each other.

[5] Privacy International (2019) Old Law, New Tech, and Continued Opacity: Policy Scotland's Use of Mobile Phone Extraction.

Collection of Personal Data is not always a Voluntary Action

The DPIA states that 'The provision of personal data is never obligatory' (p.24). It argues this by confusing the personal data processed by the App and the system with the data uploaded by each user. This argument is flawed.

The main flaw in this argument is the NHSX system is designed such that identifiable personal data which relates to a device ID **is uploaded by other users about the data subject by design, not just the data subject themselves**.

Take a trivial example. Three users sit in a cafe, one on a table by themselves and two on a table next to each other. The one on a table by themselves later uploads data after declaring/testing positive. That data reveals that the two users in the cafe who did not upload data were

1. Present in the same place as each other
2. Present at the same time as each other
3. Present at the same place and same time as the uploading user

Therefore, a third party uploads personal data describing a connection between two other people, **without the specific consent of the users concerned**. This does not happen in e.g. a decentralised system, as users never upload information about other people.

The DPIA half admits this, with the confusing paragraph:

> > The provision of personal data is never obligatory. However, if user A and user B have a proximity encounter, and user A chooses to upload data about their encounter, then the provision of that data was not speciically [sic] voluntary for user B. (p.24)

- **The document (and associated public messaging) must be changed throughout to reflect the fact that it is not the case that personal data about a user is only uploaded with a user's permission, as *other people* upload data revealing a user's social interactions.**

Systematic Monitoring

The DPIA states:

> > Will there be a systematic monitoring of a publicly accessible area on a large scale (e.g. CCTV)? **No**

The answer to this question should be yes, although it does not have significant legal consequence given the other risks this processing poses already trigger measures such as an obligatory DPIA. The large area in question is either the entire jurisdiction of at least NHS England (and other devolved nations of the UK that may choose to use this app), and/or the postcode regions attached to Sonar IDs. This entire system effectively turns every individual into a sensor and a broadcaster, monitoring every other person with the app in the area. It is designed to systematically monitor and shape behaviour in publicly and privately accessible areas.

This has already been acknowledged by Matthew Gould, NHSX CEO, who stated that the system is designed to identify 'hotspots' based on the interaction data.[6]

- **The DPIA must acknowledge the NHSX App systematically monitors publicly accessible spaces**.

Users deprived of data protection rights

The DPIA states that:

> > Is there the risk that data subjects might be deprived of their rights and freedoms or prevented from exercising control over their personal data?: **No** (p.12)

This response contradicts the DPIA and oral evidence given to Parliament, as this note will now explore.

Deprivation of the right of erasure (GDPR, art 17)

The NHSX App appears to deny people access to the right to erasure (also called the 'right to be forgotten') without a specified lawful reason for doing so.

NHSX CEO Matthew Gould stated in oral evidence to Parliament:

> > The data can be deleted as long it is on your own device. Once it is uploaded, it becomes enmeshed in wider data, and the technical difficulties of deleting it at that point become tricky.[7]

The DPIA does not properly address this issue, however it does state:

> > Users may uninstall the App from their phone at any time which will cause deletion of all the app data from the device. **This will not cascade to the Sonar backend.** [emphasis added] (p.26)

[6] Oral evidence from Matthew Gould to the Joint Committee on Human Rights, HC 265, Monday 4 May 2020.
[7] Oral evidence from Matthew Gould to the Joint Committee on Human Rights, HC 265, Monday 4 May 2020.

In combination, this appears to imply that users will be unable to delete their data, or make a request to do so. There may be a lawful basis that can be established for denying an erasure request,[8] however this is not specified.

- **The lawful basis for a blanket refusal of the right to erasure is unspecified by NHSX in this DPIA.**

It is likely that NHSX would rely on a logic of denial similar to that to which they elaborate slightly more on in relation to the right of access, which this note looks at next.

Unlawful deprivation of the right of access (GDPR, art 15)

The right of access is a critical right in data protection law. It is the tool through which users can understand how data about them is being processed, and can challenge its lawfulness. Therefore, the DPIA asks a critical question:

> > How will it be possible to provide a copy of the personal data processed about a particular individual to them (redacted as necessary) should they request access to this information?

NHSX responds:

> > This will require users to have access to their Sonar ID. With this they may be able to make a request which will be processed via the DHSC SRR process. The technical practicality of this needs to be assessed. If users do not have access to the Sonar ID, SRRs may be exempt under Article 11. (p.26)

This initially seems to read as if such a request would be possible. However earlier in the document NHSX states:

> > These codes [..] are not linked to any information that identifies the user, **and the user cannot access them**. These are [..] The Sonar ID: a code that is given by the central database when the user registers [..] [emphasis original] (p.6)

So, effectively, what has happened, is that the Sonar ID identifier that would be necessary to pull (and also, to delete) your data from the centralised database is *deliberately buried in the app, and not surfaced to the user*, with the effect that individuals are deprived of their rights. This type of practice is arguably a violation of the GDPR, art 25, *data protection by design*.[9] Data protection by design means that the rights and obligations of data protection law must be designed into the systems a data controller builds. In this case, they have been designed out.

[8] See generally Jef Ausloos, *The Right to Erasure in EU Data Protection Law* (OUP 2020).

[9] See further Michael Veale and others, '[When Data Protection by Design and Data Subject Rights Clash]' (2018) 8 International Data Privacy Law 105. doi:10/gdxthh

It could be argued that designing them out would be justified on balance if they instead promoted, equally or to a greater degree, one of the other data protection principles. However, it is not described why an individual's Sonar ID would not be able to be provided to the user, or the risks it would create. The Sonar ID is only used to index data in the central database, and so it would be useless to anyone unless they were to break into the NHSX database, which they have assured the public will be kept secure.

NHSX mentions Article 11 of the GDPR. Article 11(2) provides that:

> Where [..] the controller is able to demonstrate that it is not in a position to identify the data subject, the controller shall inform the data subject accordingly, if possible. In such cases, Articles 15 to 20 shall not apply except where the data subject, for the purpose of exercising his or her rights under those articles, provides additional information enabling his or her identification.

As NHSX notes, the identification of the individual in the system is easily verified using their Sonar ID. Furthermore, the NHSX app is in regular two-way communication with the NHSX server. The controller is therefore in a position to identify the data subject *for the purpose of exercising data rights*. These do not require their real identity to exercise, because, as the NHSX states, there is no need for the real identity to deliver the service, and therefore this data would be unnecessary to process for the purpose of providing functioning data rights, such as access or erasure. The individual is already identifiable for the purposes of transmitting the information held on the server correctly to their device (and not to any other device), or for transmitting an erasure request.

Even *if* there were good arguments to hide the Sonar ID from the user (such as security), NHSX could make the app generate a one off 'SAR ID' which would be cryptographically linked to the Sonar ID in the same way as transmitted IDs are. Unlike transmitted IDs, it would only show up in the phone, not be broadcasted. Given that the user regularly broadcasts such IDs into the public, where they can be received by anyone with a Bluetooth device, there can be no security argument that such data needs to be kept confidential.

- **The NHSX App unlawfully designs out the right to access when there is a legal obligation to design it in.**

Right to object

One right that is not mentioned in the DPIA is the right to object (GDPR, art 21).

This is crucial to mention as it *specifically applies* to the lawful basis being used by DHSC, GDPR art 6(1)(e). Individuals must have the ability to object 'on grounds relating to his or her particular situation'. From that it follows that blanket rejections are not permitted, as a data subject's specific situation must be possible for the data controller to consider and fairly appraise.

- **If the controller plans to, as with the right to erasure and the right to access, refuse all attempts at the right to object, this needs a justification in the DPIA.**

Deprivation of rights requires a legislative provision

The Government has another option, if it is keen to ensure individuals are unable to ask for a copy of their personal data, or to be able to erase it. It can enact specific legislative provisions to extend the situations for exemptions to these rights (under GDPR, art 23). The current list of these exemptions relating to health can be found in Schedule 3 of the Data Protection Act 2018.

Automated decision-making

It is important to see that NHSX agrees with the Council of Europe (the UK is a signatory of the modernised data protection Convention 108) that the NHSX app constitutes an automated decision with legal effect or similarly significant effect under Article 22 of the GDPR. Based on the DPIA, they are stating that it will be considered as a 'qualifying significant decision' under the Data Protection Act 2018 s 14(3). They also state the legal basis for this decision as Regulation 3(1) and 3(3) of the Health Service (Control of Patient Information) Regulations 2002.

A basis in law would be needed for such a decision. However, there is a problem. The Regulations cited above are suitable for authorising processing, but not suitable for authorising *automated decision-making*.

Processing under data protection legislation does not include decision-making

The decisions in question are 'decision[s] *based solely on automated processing*', following GDPR, art 22. As can be seen, processing is a separate activity here, undertaken *prior to the decision* being made. GDPR, art 22(2) is specific that it must be 'the decision', not the processing, which 'is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests'. The processing must, of course, *also* have a legal basis, but how that is achieved relates instead to GDPR, art 6.

The definition of processing within the meaning of the GDPR does **not** include decision-making, referring to the operations performed on data, rather than decisions resulting from it.[10] The reason for this has been made clear in English law by the Court of Appeal, on the basis that, if processing did extend to include decision-making, it would

[10] Specifically GDPR art 4(2) states that processing is 'any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction'.

mean that all public or private sector decision-making aided by a computer would be subject to data protection principles of fairness (while those not aided by a computer or filing system would not).[11]

Further confirmation can be found in the Data Protection Act 2018 sch 1 paras 20(2)(a) and 21(1)(c), which in its drafting clarifies that decision-making is a *purpose* for which processing is carried out for, rather than part of the processing itself.

Processing under the Health Service (Control of Patient Information) Regulations 2002 does not include decision-making

The DPIA states that the lawful basis for the solely automated, significant decision is regulation 3(1) and 3(3) of the Health Service (Control of Patient Information) Regulations 2002.

Under these Regulations, 'processing' is further re-defined to clarify that it

> *includes (in addition to the use, disclosure or obtaining of information) any operations, or set of operations, which are undertaken in order to establish or maintain databases for the purposes set out in paragraph (1), including—*
> *(a) the recording and holding of information;*
> *(b) the retrieval, alignment and combination of information;*
> *(c) the organisation, adaption or alteration of information;*
> *(d) the blocking, erasure and destruction of information.*

While the regulation provides a more specific definition of processing, it does not include decision-making in its meaning. Indeed, risk scoring and alerting an individual cannot be seen as an operation that is 'undertaken in order to establish or maintain databases'.

Furthermore, the DPIA states that it utilises Article 9(2)(h) and Article 9(2)(i) of the GDPR as grounds to overcome the prohibition on special category data.

However as the DPIA confirms, a decision to assess a person as potentially infected and/or to send guidance to their contacts, is a solely automated decision under art 22 of the GDPR. This has real consequences as Article 22(4) says that such a solely automated decision cannot be made re sensitive personal data (e.g. health data) except in highly limited circumstances ie explicit consent (art 9(1)(a) or substantial public interest (art 9(1)(g)). Unfortunately NHSX asserts in the DPIA that the ground for processing sensitive personal data in the app is art 9(2)(h) and (i). This is seriously misleading. NHSX needs to rely on Article 9(2)(g) by asserting that the mandate in the Health Service (Control of Patient Information) Regulations (COPI) is also a justification for a claim of "substantial public interest" as lawful basis, and amending the DPIA . However COPI does still *not* lay down 'specific' safeguards

[11] *Johnson v Medical Defence Union* [2007] EWCA Civ 262 [44]–[48].

relating to automated decision-taking, and so they do not appear appropriate to authorise an Article 22 decision.

As a consequence, COPI, as they currently stand, do not constitute a lawful basis for an automated decision under the Data Protection Act 2018 s 14 and new legislation (albeit brief) is probably necessary.

- **The DPIA does not set out a valid lawful basis for the solely automated, significant decision-making it correctly identifies as occurring.**

Logic of automated decisions

There is an obligation under GDPR art 13 to describe the logic in automated decision-making. NHSX appears to have provided this in an embedded PDF inside the DPIA. This cannot be opened, as such a method of embedding only works within Microsoft Word files. Furthermore, while it is important that they have done that, it is more important that such a document be both accessible at all as well as provided alongside the GDPR art 13 information in the privacy policy.

- **The information contained in the document embedded in the DPIA describing the logic of automated decisions must be provided under GDPR, article 13.**

Risks

The published DPIA is incomplete, as the application and platform risks have been redacted. **Whether these have been assessed properly therefore cannot be scrutinised by the public**.
At the very least, the public should be aware of the following risks from adversaries who are *not the data controller*, which public academic work indicates will also apply to the NHSX system.[12]

1. A motivated adversary can **identify the infected people that they have been in close proximity to**. This risk is a consequence of the basic proximity tracing functionality and does not depend on any design choices or implementation details, and so applies to all proximity tracing systems, including the NHSX app. (risk to security principle)
2. An adversary can **trigger false alerts about encounters with an infected person that do not necessarily reflect real-world physical proximity**. (risk to accuracy principle, security principle). This is exacerbated by self-reporting.
3. An adversary can **disrupt the contact discovery between users** through noise injection in the radio channel. (risk to accuracy principle)

[12] The DP-3T Project (20 April 2020) *Privacy and Security Risk Evaluation of Digital Proximity Tracing Systems*.

4. An adversary **can track users based on aspects orthogonal to contact tracing** (e.g. MAC addresses) (risk to security principle).
5. An adversary that actively emits identifiers over time, such as in a vehicle in a large space using rotating accounts, **can identify locations with infected people present** (risk to security principle).

Furthermore, in the NHSX system specifically

6. Any adversary **can track a user based on the Bluetooth signal emitted by their contact tracing app for one day**. (risk to security principle). NHSX could change this to 15 minute rotation, but in the current architecture, would be unlikely to be able to align it with MAC rotation, meaning linkage could be made by comparing the overlap with MAC addresses to allow for longer tracking.

These issues could, for example, cause panic, social stigmatisation, adverse health outcomes, or could exacerbate commercial or other forms of tracking.

Article 36 Prior Notification

None of the above risks, except for risk 6 which the controller has chosen not to remove, can be removed in their entirety by NHSX. Identification of individuals' health status without permission is an especially high risk. This is possible by using a modified version of the app, creating many accounts (which there is no solid technical barrier to doing), and cycling through them as the attacker meets different people and spends time with them. Ensuring that each account is only used when significantly proximate to one person, and then disabling the account, allows the user to then wait to see if they get a risk event. If they did, then they know the individual they were proximate to tested positive or self-declared results. Accounts can be made to look active by making it appear they are regularly close to other accounts the attacker controls, preventing detection as an anomalous short term account.

This is just one example of a risk which is inherent to Bluetooth contact tracing systems (in this case, a Sybil attack). It is clearly a high risk because, as the DPIA states, this information is *confidential health information* which the system risks exposure to by any person.

Article 36 of the GDPR states that in cases where high risks cannot be mitigated in full, ***before*** proceeding with processing, the Information Commissioner must be consulted. According to this DPIA, that does not appear to have occurred. The DPIA states the Information Commissioner was to be 'briefed' rather than consulted (p.22).

- **The Information Commissioner must be consulted prior to processing within the meaning of GDPR, art 36.**

The Privacy and Electronic Communications Regulations

Regulation 6 of The Privacy and Electronic Communications (EC Directive) Regulations 2003 (PECR) provides that there must be a legal basis for an individual to gain access to information in the terminal device of another user. This would include the Bluetooth handshake mechanism envisaged by NHSX in this app.

This also specifically relates to the trackers discovered inside the NHSX app.[13] These trackers are not necessary for the service explicitly requested by the end user. These need to be declared explicitly to the user before installation, as the ICO notes.[14]

- **The DPIA should explain how the The Privacy and Electronic Communications Regulations are complied with, both in relation to Bluetooth usage and in relation to embedded trackers.**

[13] Privacy International, UK government Covid tracking app: what you need to know! (7 May 2020).
[14] https://ico.org.uk/for-organisations/guide-to-pecr/cookies-and-similar-technologies/