Data Dictionary

To aid discussion of data privacy, data use ethics and probity implications, this document collects the definitions and usage of all data points used in the operation of the Sonar app and backend.

Usage is described in three forms

- As answers to end user questions about the journey they take with the app
- As a summary table of data usage scope
- In detail per data point

Journey-based user questions

- When I download the app
    - What is stored by the Apple App Store or Google Play?
        - The fact that you downloaded the app and which device you downloaded it to.
    - What is recorded in my browser history?
        - If you follow the link in the application to the explanation or advice pages on the NHS website, the address of the pages you visit will be recorded in the browser history on your phone.
- When I register
    - What is stored in the NHS database?
        - Your app user id
        - The make and model of your phone
        - A messaging service id for this app only, which the app will use to send you messages
        - Your home postal district - the part of your postcode before the space
    - What is stored on my phone by the app?
        - Your app user id
        - A messaging service id for this app only, which the app will use to send you messages
- When I have the app running
    - What is visible to Bluetooth devices near me?
        - Your temporary app user id for the day
    - What is monitored by my phone?
        - Power and data usage of the app will be monitored by your phone
- When I am close to someone else who has the app
    - What is stored on my phone by the app?
        - The other person's temporary app user for the day
        - Information about how strong the Bluetooth connection was between the two phones
    - What is stored on the other person's phone?
        - Your temporary app user id for the day
        - Information about how strong the Bluetooth connection was between the two phones
- When I update my symptoms?
    - If I have no symptoms?
        - Nothing
    - If I have symptoms?
        - What is stored in the NHS database?
            - The data stored in your phone about other phones you have been close to in the last 28 days.
            - The symptoms you have reported
            - The date your symptoms started

- When I get a notification that I have been in contact with someone who has symptoms?
    - What is recorded in my browser history?
        - If you follow the link in the application to the explanation or advice pages on the NHS website, the address of the pages you visit will be recorded in the browser history on your phone.
- If I delete the app
    - What is still stored in the NHS database?
        - Your app user id
        - The make and model of your phone
        - A messaging service id for this app only, which the app will use to send you messages
        - Your home postal district - the part of your postcode before the space
    - What is stored on the phone by the app?
        - Your app user id
        - A messaging service id for this app only, which the app will use to send you messages
- Who at the NHS will use my data and how?
    - NHS staff will use your data to plan the NHS response to COVID-19, including changes in how the app works and allocating equipment and staff around the country.
- Who does the NHS share my data with?
    - Scientists advising the NHS on how to manage COVID-19 will use your data in their mathematical models to improve planning but they will not see individual records.

Retention period

In accordance with the law, personal data will not be kept for longer than is necessary. The exact retention period for data that may be processed relating to COVID-19 for public health reasons has yet to be set (owing to the uncertain nature of COVID-19 and the impact that it may have on the public).

In light of this, we will ensure that the necessity to retain the data will be routinely reviewed by an independent authority (at least every 6 months).

There will be a research value for data selected by the NHS COVID-19 App, along with any other COVID-19 data set. Whilst the NHS COVID-19 App will ensure that information processed within the NHS COVID-19 App cannot be identified, there may be requests to process data from the app for research purposes, which may be linked with identifiable data. All such requests will be subject to further approvals and independent oversight.

Data Scope Summary

The table below shows all the data points collected or derived during the operations of the Sonar mobile application and backend processes. The list is grouped by the subject of the data point - person, device, application store user, application instance, encounter. More detailed discussion of each data point is grouped in the same way below the table.

The scopes within which each data point visible are indicated in the right-hand columns of the table

LA - Local Application - the Sonar application on the device
PA - Proximate Application - the Sonar application on any device with range for Bluetooth Low Energy (BLE) discovery
SB - Sonar Backend - the central application services and databases supporting the operation of the Sonar application
TP - Technology Partners - providers of systems which are used to distribute the application, verify application installation, send messages to the application, etc
CR - Covid Response - Other organisations in the NHS or outside working on the COVID response, subject to case-by-case analysis of need
3P -Third Parties - Other organisations or researchers not involved in COVID response

Captured Values							
Person							
Label	Description	LA	PA	SB	TP	CR	3P
COVID-19 Self-Diagnosis	Whether this person has self diagnosed as suffering from COVID-19	✓		✓		✓	
Symptoms	Fever and/or persistent cough	✓					
Onset of Symptoms	Date the symptoms above began	✓		✓		✓	
Postal District	First portion of Post Code	✓		✓		✓	**X**
Device							
Label	Description	LA	PA	SB	TP	CR	3P
Unique ID	Hardware-level unique identifiers for device	---	---	---	✓	---	---
Make and Model	Manufacturer and model code of device	✓		✓	✓		
Application Distributor User							
Label	Description	LA	PA	SB	TP	CR	3P
User ID	Vendor-specific user identifier issued by app store	▫[1]			✓		
Application Instance							
Label	Description	LA	PA	SB	TP	CR	3P
Messaging ID	Identifier issued by the messaging service	✓		✓	✓		**X**
Sonar ID	Identifier issued by Sonar to uniquely identify participating devices independently of any PII	✓		✓		✳[2]	**X**
Transmitted ID	Encrypted identifier used to identify participating devices to each other when in range	✓	✓	✓			**X**

[1] This field is available to our Mobile application on Android and iOS but we do not use it
[2] Exports which require linking data for devices will include one-time tokenisations of each Sonar ID

Encounter							
Label	Description	LA	PA	SB	TP	CR	3P
Transmitted ID	Encrypted identifier used to identify participating devices to each other when in range	✓	✓	✓			**X**
Timestamp	When the devices communicated	✓	✓	✓		✓	
Radio Signal Information	Transmit Power, Radio Signal Strength Indicator for the Bluetooth interaction	✓	✓	✓		✓	
Connection with other NHS systems							
Label	Description	LA	PA	SB	TP	CR	3P
Reference Code	Temporary identifier used for Sonar to receive information about a user from other NHS systems	✓		✓		✓	
Derived Values							
Label	Description	LA	PA	SB	TP	CR	3P
Distance	How close two devices were to each other			✓		✓	
Encounter Duration	The period for which the devices were close to each other			✓		✓	
Mobile Application Usage Metrics to be collected and processed during Isle of Wight limited release only							
Label	Description	LA	PA	SB	TP	CR	3P
AppCentre Required Data	Minimum data required for Microsoft AppCentre to operate	▫			✓		**X**
Completion of Postal District field	Whether the user entered their postal district data into the app				✓		**X**
Completion of Permissions process at installation	Whether the user completed the process granting required permissions for the app.				✓		**X**
Mobile Operating System	The operating system running on the device	✓			✓		**X**
Failure to get Firebase Token	Whether the app successfully obtained a Messaging ID from Firebase				✓		**X**

Failure of Device Registration	Whether the app successfully obtained a Sonar ID from the backend				✓		**X**
Successful Device Validation	Whether the app received a validation message from the backend				✓		**X**
Number of contact events recorded	Count over a calendar day of BLE connections established by the device	✓			✓		**X**
Number of contact requests received	Count over a calendar day BLE connections established to the device	✓			✓		**X**

Person

COVID-19 Infection Status

Whether the user of the app has self-diagnosed, been referred for testing by a clinician, or been diagnosed through testing as having contracted COVID-19.

Capture
- Submission by app user

Processing
- Identification of users to notify

Basis for Processing
- Infection status is a key data point required to achieve the project's stated intent of notifying people who have come into contact with someone infected with COVID-19.

Continuity
- Point in time, going forward

Symptoms

The symptoms a user was experiencing which led them to self diagnosis

Capture
- Submission by app user

Processing
- Used by the local app to decide on whether or not to send encounter information

Basis for Processing
- Infection status is a key data point required to achieve the project's stated intent of notifying people who have come into contact with someone infected with COVID-19.

Continuity
- Point in time, going forward

Date of Onset of Symptoms

When the user began experiencing the symptoms which led them to self diagnosis

Capture
- Submission by app user

Processing
- Used to derive risk scores for reported encounters

Basis for Processing
- The date of onset of a potential infection source's symptoms is a factor in deciding which users should be notified of having been in contact with someone who has contracted COVID-19

Continuity
- Point in time, going forward

Postal District

The first portion (up to the space) of the users' home address postcode. This is considered to be a pseudo-anonymous data item for data protection purposes.

Capture
- Entered by app user at Registration and forwarded to the Sonar backend systems during registration

Processing
- Held in the backend and associated with a sonar ID.

Basis for Processing
- NHS operations planning including availability of equipment, people and supplies in hotspots and promotion of app uptake in under-penetrated areas
- Input to social mixing estimates for epidemiological modelling

Continuity
- Potentially permanent

Device

Data about the device running the Sonar app is available to the app through API calls. Some of this information is also recorded and made available in aggregate by the application distribution channels.

Unique Identifiers

Identifiers for interacting with the mobile networks and internal hardware identifiers. Manufacturers and mobile OS providers, particularly Apple, have gone to some lengths to prevent third-party access to this identifying information.

Capture

- Not captured

Processing

- Not processed

Basis for Processing

- Given the potential for this data's use in linking application data to other data sources, recording or storing it is best avoided.

Continuity

- Permanent

Make and Model

Manufacturer string and model code string for the device as reported to the application by the operating system.

Capture

- Reported to the Sonar backend by the app during registration

Processing

- Used to interpret signal strength data to classify encounters

Basis for Processing

- Estimating distance between devices from signal strength reported during BLe interactions varies by device.

Continuity

- Permanent

Application Distributor User

Platform-specific or vendor-specific curated repositories of apps (Apple's App Store, Google Play Store, Samsung's Galaxy Store, etc) are the dominant source of apps for installation on mobile devices. All these storefronts require user registration and record device and app info for users. Developers are also given some aggregated statistics on devices which have installed an app.

Vendor-specific user identifier

To allow applications developers/vendors to identify users across their apps without compromising privacy through cross-vendor data sharing, Application Distributors provide IDs generated for each vendor. For example `identifierForVendor` on iOS or `ANDROID_ID` on Android. If all applications from a developer are deleted from a device, the identifier may be regenerated when an application from the developer is first installed again.

Due to its possible use as a key for cross-linking data from Sonar to data from other applications published by the NHS developer account, this identifier is not and should never be sent to the Sonar backend servers.

Capture

- Not captured

Application Instance

Messaging ID

Capture

- Retrieved by the application from the messaging provider during registration and sent to the Sonar backend. This is considered to be a pseudo-anonymous data item for data protection purposes.

Processing

- Sent to the massaging provider by the Sonar backend as part of sending notifications and reminders

Basis for Processing

- Required to support the sending of notification messages

Continuity

- Lifetime of the application instance

Sonar ID

Capture

- Generated by the Sonar backend during registration. This is considered to be a pseudo-anonymous data item for data protection purposes.

Processing

- Stored on the device, used to generate Transmitted ID during periodic rotation
- Derived on the Sonar Backend through decryption of the Transmitted ID
- Hashed with an extract-specific salt value when extracted for analysis

Basis for Processing

- Used to identify devices during risk scoring and messaging

Continuity

- Lifetime of the application instance

Transmitted ID

Identifier used to identify participating devices to each other when in range. This is considered to be a pseudo-anonymous data item for data protection purposes.

Capture

- Generated by the device daily (currently, can be changed).

    - Compound identifier consisting of validity start and end times, the Sonar ID, and an identifier for the instance of the Sonar application which the app is bound to
    - Instance identifiers are currently ISO 3166 country codes.
    - Encrypted with the public key for the Sonar instance which the app is bound to
    - For further info, please see the Design document on token rotation.
- Captured by proximate devices.

Processing

- Stored on the device and sent to other devices in range via Bluetooth.

Basis for Processing

- Used to identify devices to each other during encounters

Continuity

- Regenerated as required (Currently daily)

Encounter

Transmitted ID

Identifier used to identify participating devices to each other when in range. This is considered to be a pseudo-anonymous data item for data protection purposes.

Capture

- Read from the remote device when a BLE connection is established.

Processing

- Received from devices in range, stored in the local device, sent to the Sonar backend when triggered by the user
- Decrypted on the Sonar backend using the instances private key, instances are currently expected to be operated per country
- This token may be retrieved by any Bluetooth Low Energy device in range

Basis for Processing

- Used to identify devices to each other during encounters

Continuity

- Regenerated as required

Timestamp

Capture

- On receipt by the app of a BLE advertisement or periodically for active connections

Processing

- Used to derive encounter duration

Basis for Processing

- Durations of encounters are used to decide which users should be notified of having been in contact with someone who has contracted COVID-19

Continuity

- Once-off

Radio Signal Data

Radio Signal Strength Indicator is a value for the strength of the BLE signal reported to the application by the operating system, along with the reported Transmit Power and Receive Power for the BLE interaction

Capture

- On receipt by the app of a BLE advertisement or when requested by the app for an active connection

Processing

- Used to derive proximity range estimates for encounters.

Basis for Processing

- Range estimates for encounters are used to decide which users should be notified of having been in contact with someone who has contracted COVID-19

Continuity

- Once-off

Connection with other NHS systems

Reference Code

Temporary identifier used for Sonar to receive information about a user from other NHS systems such as virological testing result systems

Capture

- Generated on the Sonar Backend when requested by the user

Processing

- Displayed to user or sent to participating peer system as required
- Deleted after corresponding data is received from participating peer system or on expiry

Basis for Processing

- Enabling data acquisition from other NHS systems without sharing personal identifiers

Continuity

- From request to peer system response or expiry

Application Usage Journey

During the limited release in the Isle of Wight, additional data will be collected about the app's performance and user interactions with the app. This data will be captured through Microsoft AppCentre and will not include the SonarID or any other identifying information. This is necessarily held on Microsoft servers away from our platform, as some of the metrics are for the inability to contact our service.

AppCentre Required Data

Minimum data required for Microsoft AppCentre to operate, as documented under the 'required' subheaders on https://docs.microsoft.com/en-us/appcenter/sdk/data-collected. For convenience these are

- Application Secret (identifying the app to AppCentre)
- Installation ID (identifying the app instance to AppCentre)
- Application properties - Version, Build
- SDK properties - Name, Version
- Operating System properties - Name, Version
- Device Configuration - Language and Country Code, Time Zone Offset

Capture
- Captured whenever a metric event in the list below is sent to AppCenter

Processing
- Uploaded to Microsoft AppCentre, summarised for visual analysis

Basis for Processing
- Requirement for gathering data on the Application Usage Journey during limited release on the Isle of Wight

Continuity
- Potentially for the lifetime of the device

Completion of Postal District field

Whether the user entered their postal district data into the app

Capture
- When a user clicks Next after entering their postal district (the portion of the postal code before the space, for instance 'DE1')

Processing
- Uploaded to Microsoft AppCentre, summarised for visual analysis

Basis for Processing
- To determine is asking for a partial postcode adversely affects adoption of the app

Continuity
- Point in time

Completion of Permissions process at installation

Whether the user completed the process granting required permissions for the app.

- On iOS
    - Bluetooth

        - Notification
    - On Android
        - Bluetooth
        - Location - required for Android to allow Bluetooth Low Energy beacon operations

Capture
- When the user grants the required permissions

Processing
- Uploaded to Microsoft AppCentre

Basis for Processing
- To determine is the required permissions adversely affect adoption of the app

Continuity
- Point in time

Mobile Operating System

The operating system in use on the device

Capture
- Captured for all events by Microsoft AppCenter, named specifically due its direct use

Processing
- Uploaded to Microsoft AppCentre, used to interpret permission completion data above

Basis for Processing
- Used to determine which of the two mobile applications has a problem, success, or to segment other statistics in this section for future improvement of the application and backend service

Continuity
- Lifetime of the device

Failure to get Firebase Token

Whether the app successfully obtained a Messaging ID from Firebase

Capture
- On failure of the request to Google Firebase for an application token

Processing
- Uploaded to Microsoft AppCentre

Basis for Processing
- Measuring the reliability of Firebase registration

Continuity
- Point in time

Failure of Device Registration

Whether the app successfully obtained a Sonar ID from the backend

Capture

- When the request to the backend servers fails

Processing
- Uploaded to Microsoft AppCentre

Basis for Processing
- Measuring the reliability of the registration process

Continuity
- Point in time

Successful Device Validation

Whether the app received a validation message from the backend

Capture
- On receipt of a validation message and completion of application registration

Processing
- Uploaded to Microsoft AppCentre

Basis for Processing
- Measuring reliability of the device validation process

Continuity
- Point in time

Number of contact events recorded

Count over the previous calendar day of Bluetooth connections established by the device

Capture
- When a nearby device's Transmitted ID is successfully read

Processing
- Uploaded to Microsoft AppCentre after the end of each day

Basis for Processing
- To understand the validity of the Bluetooth contact tracing protocol we are using

Continuity
- Point in time

Number of contact requests received

Count over the previous calendar day of Bluetooth connections established to the device

Capture
- When a nearby device reads the local device's Transmitted ID

Processing
- Uploaded to Microsoft AppCentre after the end of each day

Basis for Processing
- To understand the validity of the Bluetooth contact tracing protocol we are using

Continuity
- Point in time