Data Protection Impact Assessment – The NHS OpenSAFELY Data Analytics Service Pilot

Document filename:	**The NHS OpenSAFELY Data Analytics Service Pilot 2025 DPIA**	
Directorate / Programme	**Transformation / Data & Analytics**	
Document Reference (previous reference number: IG20200468)		
Information Asset Owner	Michael Chapman Director of Data Access and Partnerships	Version 1.2
Author	Narissa Leyland Head of Data Governance & Assurance	

Document Management

Revision History

Version	Date	Summary of Changes
1.0	14/07/2025	Version 1.0 approved
1.1	14/01/2026	Addition of Improving Access to Psychological Therapies to the list of NHS England controlled datasets
1.2	12/03/2026	Amendment to clarify that Civil Registrations data will flow from NHS England and not from ONS. Amendment to clarify that the Civil Registrations dataset will include full date of death

Reviewers

This document must be reviewed by the following people:

Reviewer name	Title / Responsibility	Date	Version
REDACTED	Deputy Director of Data Access & Partnerships	16/01/2025	0.10
REDACTED	Assistant Director of IG	17/04/2024	0.5
REDACTED	Deputy Director Information Governance Delivery (Digital & Operations)	09/06/2025 16/01/2026	1.0 1.1
REDACTED	Data Protection Officer	08/07/2025	1.0

Approved by

This document must be approved by the following people:

Name	Title / Responsibility	Date	Version
Michael Chapman	Director of Data Access and Partnerships /IAO	15/01/2026 13/03/2026	1.1 1.2
REDACTED	Deputy Director IG Delivery (Digital & Operations)	14/07/2025 16/01/2026 13/03/2026	1.0 1.1 1.2

Document Control:

The controlled copy of this document is maintained in the NHS England corporate network. Any copies of this document held outside of that area, in whatever format (e.g., paper, email attachment), are considered to have passed out of control and should be checked for currency and validity.

Contents

Glossary of Terms

Term / Abbreviation	What it stands for/means
Pilot	Short-term enablement for the service to be operated and used for purposes that are wider than the COVID-19 Purposes and evaluate the technology
The Service	The NHS OpenSAFELY Data Analytics Service
GPSS	GP System Suppliers
TPP	The Phoenix Partnership Ltd (Leeds)
EMIS	Egton Medical Information Service Ltd
Optum	New name for EMIS GP System Supplier
Approved Users	Approved Users are: • academics, analysts and data scientists who have signed a data access agreement, who have approved projects • able to run Queries for their approved projects • able to access Aggregated Outputs and Logs generated from their Queries inside the secure environment; and request release of Aggregated Outputs following review by output checkers against statistical disclosure control rules
Intermediate Outputs	Pseudonymised patient-level datasets - Pseudonymised datasets specifically tailored to the needs of an approved project: the project datasets have substantially less detailed information than the full coded pseudonymised data as they are a subset of the complete pseudonymised GP and NHS England patient dataset
Aggregated Outputs	Anonymous datasets - Data that does not describe attributes (for example health information) about any specific patient; it is summary data, such as the number of patients with diabetes who were vaccinated, or the number of patients with asthma who were admitted to hospital, or the counts of a particular investigation or medication. Aggregated Outputs describe particular attributes related to groups of individuals, or events.
Pseudonymised and further de-identified dataset	The dataset has been pseudonymised for direct identifiers and further de-identified with respect to most indirect identifiers, ensuring minimal risk of re-identification while maintaining the possibility of secure linkage via a pseudonym

1. Purpose of this document

A Data Protection Impact Assessment (DPIA) is a useful tool to help NHS England demonstrate how we comply with data protection law.

DPIAs are also a legal requirement where the processing of personal data is *"likely to result in a high risk to the rights and freedoms of individuals".* If you are unsure whether a DPIA is necessary, you should complete a DPIA screening questionnaire to assess whether the processing you are carrying out is regarded as high risk.

By completing a DPIA you can systematically analyse your processing to demonstrate how you will comply with data protection law and in doing so identify and minimise data protection risks.

2. Background

From 1 February 2023, NHS England has assumed responsibility for all activities previously undertaken by NHS Digital. This includes running the vital national IT systems which support health and adult social care, as well as the collection, analysis, publication, and dissemination of data generated by health and social care services. The statutory functions of NHS Digital transferred to NHS England under the Health and Social Care Information Centre (Transfer of Functions, Abolition and Transitional Provisions) Regulations 2023.

The Health and Social Care Act 2012 ('the Act') gives NHS England statutory powers, under section 259(1)(a), to require data from health or social care bodies, or organisations that provide publicly funded health or adult social care in England, that it considers necessary or expedient to have to carry out its functions under chapter 9 of the Act. This includes where it has been directed to establish an information system by the Secretary of State for Health and Social Care.

The data, as specified in the associated Requirement Specification and Data Provision Notice (DPN), is required to carry out its functions as conferred on it by The NHS OpenSAFELY Data Analytics Service Pilot Directions 2025, through the OpenSAFELY technology, until 31 March 2027. Therefore, organisations that are in scope are legally required, under section 259(5) of the Act, to provide the data in accordance with the Data Provision Notice.

The NHS OpenSAFELY Data Analytics Service Pilot (referred to as the **Service**) provides a secure analytics service for **Approved Users** to access pseudonymised GP and NHS England patient data for **Approved Projects**. The Service may only be operated for the following purposes:

- clinical audit[1];
- service evaluation[2];
- health surveillance[3];

[1] As defined in the https://www.hra-decisiontools.org.uk/research/docs/DefiningResearchTable_Oct2022.pdf

[2] As defined in the https://www.hra-decisiontools.org.uk/research/docs/DefiningResearchTable_Oct2022.pdf

[3] As defined in the https://www.hra-decisiontools.org.uk/research/docs/DefiningResearchTable_Oct2022.pdf

- research[4];
- evaluation of the Service; and
- health and social care policy, planning and commissioning purposes and public health purposes, where agreed on a project specific basis by or on behalf of:
    - the Department of Health and Social Care,
    - NHS England,
    - and a nominated representative of each of the Royal College of General Practitioners and the British Medical Association on behalf of the Joint GP IT Committee

The Service is operated by NHS England (as the **data controller**) with the Bennett Institute for Applied Data Science (University of Oxford) (as **data processor**) and The Phoenix Partnership (Leeds) Ltd (**TPP**), or Optum (formerly Egton Medical Information Systems Ltd (**EMIS**)) (the **GP System Suppliers;** also as **data processors**). The data protection roles are explored further in this document.

For avoidance of doubt, the COVID-19 Service will continue to operate independently of this Pilot and is out of scope of the NHS OpenSAFELY Data Analytics Service Pilot Direction, Requirements Specification and this DPIA. However, the Pilot and the COVID-19 Service will run in parallel and in practice will operate as one service.

The Service uses OpenSAFELY open-source software tools (**OpenSAFELY Platform**), a Trusted Research Environment, which was developed by the Bennett Institute in collaboration with the Electronic Health Record (EHR) research group at the London School of Hygiene and Tropical Medicine, NHS England, and the GP System Suppliers (GPSS). The Service uses the OpenSAFELY Platform to run project analysis code on pseudonymised GP and pseudonymised NHS England patient data which is held within TPP or Optum (formerly EMIS) systems.

The Service is designed to keep patient data confidential and enforce on Approved Users the requirement to write into computer code (for public sharing) exactly what subset of patient data they require for each approved project. Users write analysis code away from the patient data and test it on dummy data. The Service then automates the running of code (the **Queries**) across the full coded pseudonymised GP data and NHS England pseudonymised patient data linked to generate intermediate pseudonymised datasets specifically tailored to the needs of the project. These study (or project) datasets (**Intermediate Outputs[5]**), have substantially less detailed information than the full coded pseudonymised data as they are a subset of the complete pseudonymised GP and NHS England patient dataset made available by the GPSS (example in Appendix 2). Further queries are run to generate aggregated outputs (the **Aggregated Outputs[6]**), which includes logs to help identify and fix errors in the analysis code.

Approved Users can only access the Aggregated Outputs generated by their Query, and review logs (including those to help identify and fix errors in their analysis code) inside the GPSS's secure environment. Logs may sometimes contain small amounts of highly refined and pseudonymised data about a small and arbitrarily selected subset of the population to assist researchers with their study code review. A combination of technical and process controls manages the risk of log files disclosing confidential information, which in most

[4] As defined in paragraph 3.1 pf the UK Policy Framework for Health and Social Care Research - Health Research Authority

[5] Indeterminate Outputs further explanation - Pseudonymised datasets specifically tailored to the needs of an approved project: the project datasets have substantially less detailed information than the full coded pseudonymised GP data

[6] Aggregated Outputs further explanation - Data that does not describe attributes (for example health information) about any specific patient; it is summary data, such as the number of patients with diabetes who were vaccinated, or the number of patients with asthma who were admitted to hospital, or the counts of a particular investigation or medication. Aggregated Outputs describe particular attributes related to groups of individuals, or events.

circumstances will make the data anonymous in the hands of the Approved User within the platform.

Approved Users only have access to the Aggregated Outputs and Log files for their project to enable them to review these, correct any errors and check against statistical disclosure control requirements, prior to requesting release of Aggregated Outputs from the GP System Supplier's secure environment.

Aggregated Outputs are only released outside the GP System Suppliers' secure environment after trained output-checkers working on behalf of NHS England have further reviewed them to ensure disclosure controls have been applied and the data is anonymous.

If an Approved User accesses logs that contain data that could be considered a breach of confidentiality, this must be reported through established channels identified via OpenSAFELY - https://docs.opensafely.org/outputs/viewing-released-files/#reporting-a-data-breach, and NHS England - https://digital.nhs.uk/services/data-services-for-commissioners/incident-and-service-request-process.

All actions by all Approved Users in the Service are logged in public, in real-time and all Queries are logged and published (https://jobs.opensafely.org/). No record level GP or NHS England data leaves the GP system suppliers' environment.

GP Practices are already making data available for this Service, this is a continuation of an existing service, covering all practices in England that use TPP or Optum (formerly EMIS) system.

Why is a DPIA required and what is its scope?

The UK General Data Protection Regulation (**GDPR**) requires a Data Protection Impact Assessment (DPIA) to be completed by a controller where its processing of personal data is considered to be a high risk to the rights and freedoms of individuals. In particular, the UK GDPR requires a DPIA to be carried out where there is processing of personal data relating to health on a large scale.

For the GP data, the GP Practices are the Controllers of the data within the clinical systems, with GP System Suppliers (GPSS) as Processor. The data is then processed (by the GPSS) on behalf of the GP Practice as Controller, to make a pseudonymised copy of only the coded patient data that continues to be stored within the GPSS's secure boundary. The GP Practice remains the Controller of the data at this point.

Separately, NHS England ingests other (non-GP controlled) pseudonymised data (for which NHS England is the Controller) within the secure boundaries of the GPSS. The GPSS are acting as Processors to NHS England for this activity. Details on those data sets can be found in Appendix 1, with additional technical details on the following website - Data Sources - OpenSAFELY documentation. [Placeholder – NHS England website]

NHS England then sends a Query to the GPSS, executed via the OpenSAFELY system, and the GPSS runs that query on both the GP-Controlled data and the NHS England-Controlled data and returns an NHSE-Controlled Output (Intermediate Outputs or Aggregated Outputs as described above) to NHS England (this Output remains within the secure boundaries of the GPSS, who act as Processors for NHSE for this activity).

The collection and processing of this data by NHS England requires a DPIA.

The processing of the pseudonymised data in each data set is carried out by the **Service** to support **the Purposes.**

Therefore, this document has been prepared by NHS England, as a DPIA to satisfy its own compliance requirements as a Controller of personal data, as described above.

This document should be read in conjunction with the DPIA Guidance and DPIA Screening Questionnaire (see below):

DPIA Screening Questions:

Will the processing involve a large amount of personal data (including pseudonymised personal data) and affect a large number of data subjects?	YES. The process will involve processing pseudonymised GP data (for patients in England whose GP Practice uses TPP/Oputm (formerly EMIS) as a System Supplier) with a range of other related pseudonymised datasets (see Appendix 1, with additional technical details on the following website: Data Sources - OpenSAFELY documentation) [Placeholder – NHSE website]
Will the project involve the use of a new technology(ies)?	NO.
Is there the risk that the processing may give rise to discrimination, identity theft or fraud, financial loss, damage to the reputation, loss of confidentiality of personal data protected by professional secrecy (e.g., health records), unauthorised reversal of pseudonymisation[3], or any other significant economic or social disadvantage?	No, because unauthorised reversal of pseudonymisation is not possible by the Approved User. Only NHS England, the GPSS's and the external data providers technical staff have access to the encryption keys. No system administrators and OpenSAFELY platform developers have access to the encryptions keys. The access to the keys is rigorously controlled by NHS England to only those who need it. Furthermore, once TPP prepare the datasets for analysis, they replace the pseudonym with another pseudonym known only to them (for which only TPP hold the matching table). Optum retains the original pseudonym. This significantly reduces the risk of re-identification and protects the patients.
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.

Will the processing of personal data occur without informing the individual of the processing?	NO Controllers of personal data have obligations to make patients aware of any processing of their data. NHS England is working with the GP profession to support a coherent approach to transparency for the public, but patients will not be aware of specific processing activity on their own data as the information is pseudonymised for direct identifiers and further de-identified with respect to most indirect identifiers, ensuring minimal risk of re-identification while maintaining the possibility of secure linkage via pseudonyms Transparency materials will be available on the NHS England website and will be provided to GP practices to publish for their patients; in addition the code executed by users of the OpenSAFELY platform is logged in full and in public at (https://jobs.opensafely.org/)
Will there be processing of genetic data, data concerning health or data concerning sex life?	YES - health data. NO - genetic data. There may be genetic diagnostic codes, for example stating that someone has a disease with a genetic component (e.g. “Down's Syndrome”) but there is no access to an individuals’ complex genome or actual gene sequencing information as these are not typically stored as structured or coded data in GP records; there will be coded data items on sexual orientation and other sexually transmitted diseases shared by the patient or laboratories, or from correspondence received from sexual health services which are subsequently coded into the GP record; but at no point will any sexual health clinical data be automatically included unless it is recorded directly by the GP.
Are the data to be processed revealing racial or ethnic origin, biometric data, political opinions, religion or philosophical beliefs, or trade union membership?	YES - some of these are recorded by GPs in the patient record. Ethnic origin data will be required for analysis purposes and will already be present in the GP records and in other health data

	sources linked via OpenSAFELY and controlled by NHS England.
Will there be processing of data concerning criminal convictions and offences or related security measures?	YES – Potentially as SNOMED codes, not free text, where a GP has (infrequently) chosen to capture such coded information relating to this as part of the health record.
Will personal data of vulnerable natural persons, in particular of children, be processed?	YES.
Will personal aspects be evaluated, in particular analysing or predicting aspects concerning performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements, in order to create or use personal profiles?	NO.
Will the project include a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person (e.g., a recruitment aptitude test which uses pre-programmed algorithms and criteria)?	NO.
Will there be a systematic monitoring of a publicly accessible area on a large scale (e.g., CCTV)?	NO.
Will the processing include any data matching e.g., the combining, comparing, or matching personal data obtained from multiple sources?	YES. Data from the datasets listed in Appendix 1, with additional technical details on the following website - Data Sources - OpenSAFELY documentation will be linkable to

	pseudonymised GP records held in the GPSS's systems.
Will the processing include any tracking e.g., processing which involves tracking an individual's geolocation or behaviour, including but not limited to the online environment?	NO.
Will the processing include any denial of service e.g., decisions about an individual's access to a product, service, opportunity, or benefit which is based to any extent on automated decision-making (including profiling) or involves the processing of special category data?	NO.

If the answer to any of the above questions is Y, please complete the rest of the form. If all of the screening questions are answered N, the local IG team must still sign off the DPIA.

3. Consultation with Stakeholders

NHS England has a requirement under section 258 of the Health and Social Care Act 2012, to ensure consultation has occurred with at least the following persons and groups:

- The person who gave the direction or made the request,
- Representatives of other persons considered likely to use the information to which the direction or request relates,
- Representatives of persons from whom any information will be collected,
- Other persons considered appropriate,

In developing the Direction and Requirements Specification, the Department of Health and Social Care and NHS England have consulted the following organisations:

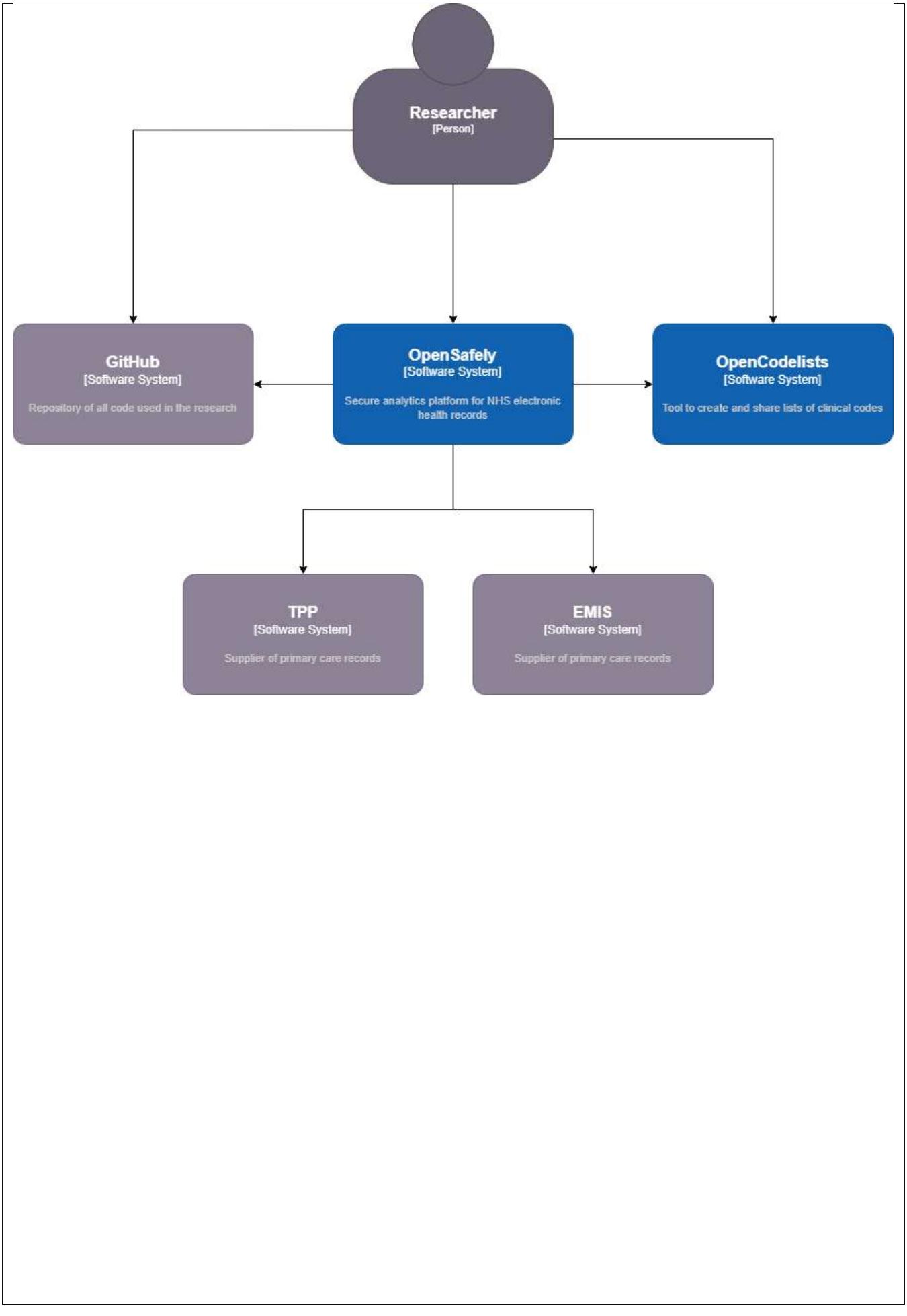
- GP Representatives including from:
    - The British Medical Association
    - The Royal College of Practitioners (RCGP)
- NHS England's Advisory Group for Data (AGD)
- Citizen Juries
- UseMyData
- The National Data Guardian for Health and Social Care
- The Data Alliance Partnership Board (DAPB)

- The Phoenix Partnership Ltd (Leeds) (TPP)
- Optum (formerly Egton Medical Information Service Ltd (EMIS))
- NHS England's OpenSAFELY Oversight Board
    - Organisational Members
        - The Bennett Institute
        - Wellcome Trust
        - Open Data Institute
        - The London School of Hygiene & Tropical Medicine (LSHTM)
        - NHS England
        - RCGP
        - BMA
        - MedConfidential
        - EMIS
        - TPP
        - Association of Professional Healthcare Analysts (AphA CIC)

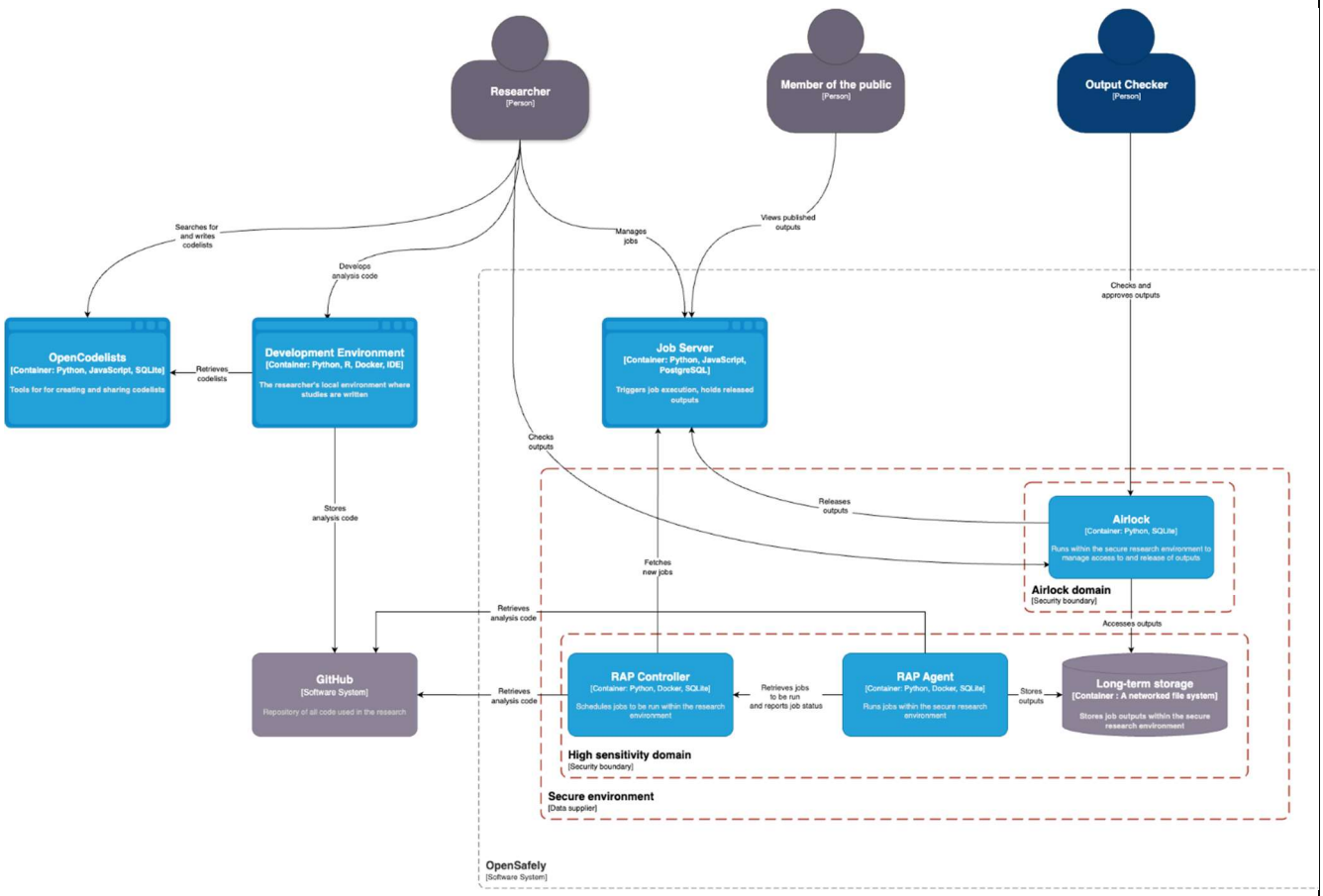
4. Data Flow Diagram

As per the details on the OpenSAFELY website, https://docs.opensafely.org/technical-architecture/, the data flow diagram is as follows:

High-level diagram of the interactions of the researcher and the tooling:


Researcher
[Person]
GitHub
[Software System]
Repository of all code used in the research
OpenSafely
[Software System]
Secure analytics platform for NHS electronic health records
OpenCodelists
[Software System]
Tool to create and share lists of clinical codes
TPP
[Software System]
Supplier of primary care records
EMIS
[Software System]
Supplier of primary care records

Detailed Technical Architecture:



The Service has been built with a focus on patient privacy and protections and, as such, aims to mitigate risk through the use of its tools and design.

“Pseudonymisation” is a widely used process for protecting patients’ privacy whereby explicit identifiers, such as names, addresses, and dates of birth, are removed from patients’ medical records before they are used or shared. Pseudonymisation creates an artificial code for each patient (a pseudonym), which allows that patient record to be linked to other records for the same patient, but without easily identifying who the individual is. Pseudonymised data is treated as Personal Data, as it is technically possible to re-identify patients from pseudonymised data; but the OpenSAFELY platform and processes use other technical controls and measures to reduce the likelihood. In addition, the GP data is not only pseudonymised for direct identifiers and but also further de-identified with respect to most indirect identifiers, ensuring minimal risk of re-identification while maintaining the possibility of secure linkage via pseudonyms. We have described this process as **pseudonymisation and further de-identification**.

Specifically for the GP Data, the following direct and indirect identifiers are removed by the GPSS in the secure GPSS environments:

- Removal of associational identifiers: Mobile phone numbers, email addresses, telephone numbers, hardware and software unique identifiers, IP addresses.
- Removal of transactional unique identifiers: all unique booking reference numbers for appointments, contacts, and referrals.
- Removal of functional unique identifiers: Titles, forenames, middle names, surnames, full dates of birth, house name, house number, street, full postcode.
- Removal of narrative text (commonly referred to as ‘free text’ data: All free text on patient records is removed. In line with other UK primary care research database

permissions, the dosage and quantity fields on prescribed medication are retained, but any script notes are removed.
- Removal of additional unstructured context: scanned images, medical drawings, letters, and all other record attachments.
- Derived data items and removal of exact original values: date of birth (MM/CCYY), partial postcodes at sector level, indices of multiple deprivation, the rurality-urban classification, geographic super-output area codes at each super output area level. Note – for organisations, the only geographic indicators stored are the lower super output areas and / or middle-level super output area code and the Local Authority code and STP/ICS code.
- Pseudonymisation is applied to the remaining data: Generation of strong pseudonyms using industry-standard cryptographic hash techniques with only NHS England approved holders of the keys to prevent re-identification of patients.

The pseudonymised datasets held within the GPSS's systems are then analysed using the OpenSAFELY platform.

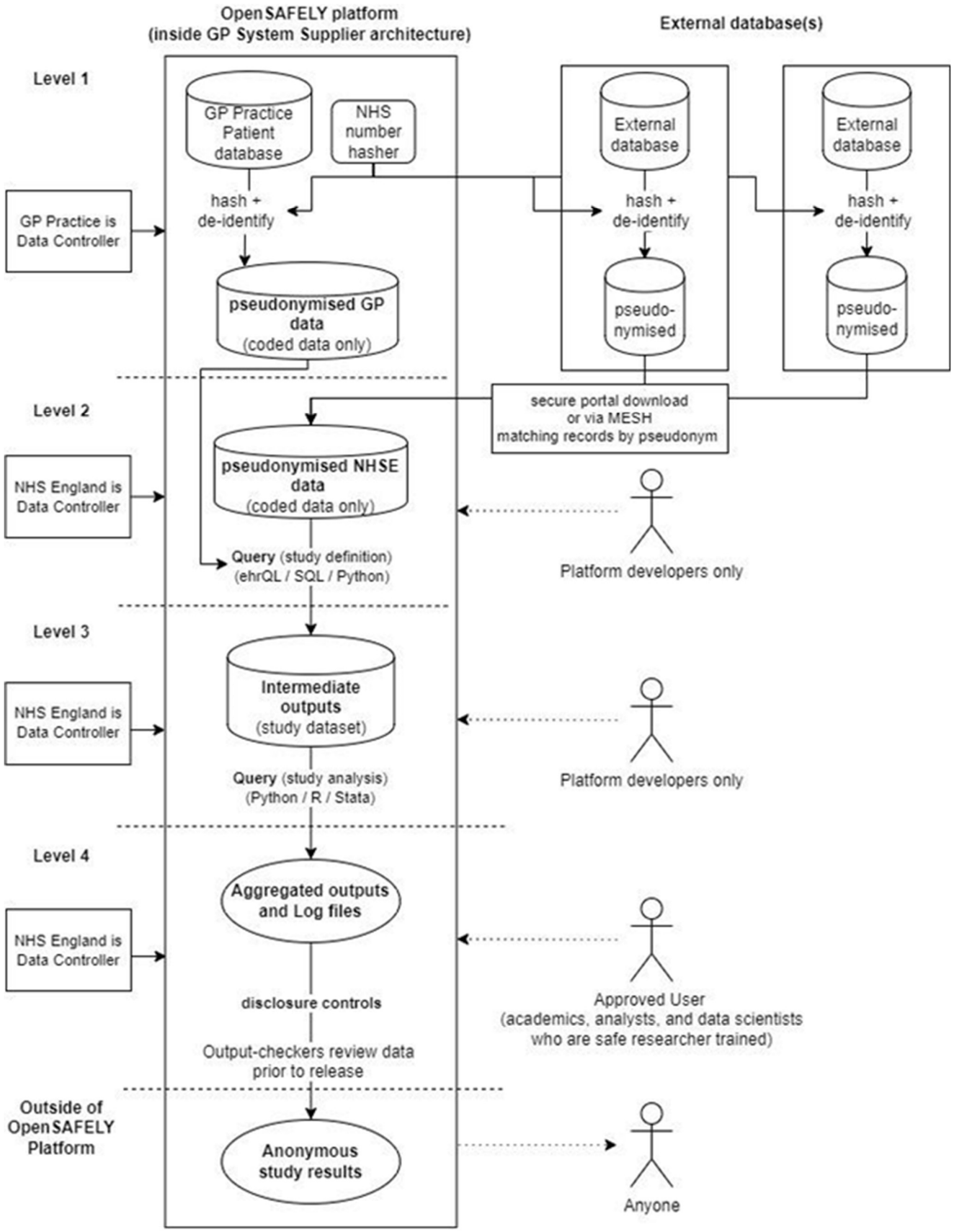
The platform allows Approved Users to design their query and run their code against randomly generated dummy data, before running the actual query as an automated request on the pseudonymised data within the Service. This eradicates the need for users to develop their analyses by interacting directly with real pseudonymised patient data and thus eradicates a key historic source of risk for projects where data is accessed at very large population scales.

The datasets (both GP-Controlled and NHS England-Controlled) are queried by the Service and the Intermediate Outputs (under NHS England Control) are the results of the Queries. These Intermediate Outputs are then subject to statistical analysis - again without the users having sight of this data - to generate Aggregated Outputs.

A limited number of Approved Users can access the Aggregated Outputs for their projects, within the Service (subject to a Data Access Agreement). The Aggregated Outputs, however, cannot be released/removed from the system or linked with other data. Prior to release from the system, there is a disclosure control process carried out to ensure the Aggregated Outputs cannot identify any individual before they are released from the system for wider sharing or publication. The disclosure control process involves the Approved Users applying disclosure controls to the results they seek to release. The results are then reviewed, cleared for release by trained output-checkers and such outputs are now considered anonymous. In addition, information relating to specific GP practices and Primary Care Networks (PCNs) is pseudonymised when the data is made available to the Service. Following agreement with the GP profession, information that could identify individual GP practices or PCNs can only be released as an aggregate output in line with a process agreed between NHS England and a representative of the Joint GP IT Committee.

As per the below diagram, the pseudonymised GP data remains under the control of GP practices (Level 1). The Service automates the running of code (the Queries) against pseudonymised GP data (level 1) and pseudonymised NHS England data (level 2), to generate intermediate pseudonymised datasets (Intermediate Outputs) in Level 3, and then final aggregated outputs (the Aggregated Outputs) in Level 4.

Detailed flow of data by level:



Type 1 Opt-Out

A software-based control operates within OpenSAFELY to uphold a patient's Type 1 Opt-out (https://www.nhs.uk/using-the-nhs/about-the-nhs/opt-out-of-sharing-your-health-records/). This Opt-out specifically means the patient wishes to prevent their GP data from being shared outside the GP practice for planning and research purposes. If a patient has registered a Type 1 Opt-out with their GP practice where they are currently registered, none of their data will be made available to an Approved Project. Specifically, none of their data held in Level 1 and Level 2 will be included in the Query, and therefore their data is not included in any Level 3 intermediate outputs (study dataset).

Technically, this control works by the GPSS storing a file that contains a list of patients (in pseudonymised form) that have exercised their right to register a Type 1 Opt-out with their currently registered GP practice in the Level 1 environment. This file is made available and used by default in the OpenSAFELY platform to ensure any pseudonym (i.e. the patient who registered a Type 1 Opt-out) does not have any of their data processed any further by approved projects. This software-based implementation approach to uphold the Type 1-Opt out has been discussed with stakeholders and has the support of the BMA and RCGP.

Data Inflows Summary table:

See Appendix 1, with additional technical details on the following website - Data Sources - OpenSAFELY documentation

Data Outflows:

All outputs are anonymised and / or aggregated.

5. Purpose of the processing

The Secretary of State has directed NHS England to collect, process and analyse data to support health functions, as set out in The NHS OpenSAFELY Data Analytics Service Pilot 2025. This enables NHS England to collect data, analyse and link data for the Purposes using data held within the secure environments within the GPSS's systems with access via the OpenSAFELY Platform.

This DPIA, published in association with The NHS OpenSAFELY Data Analytics Service Pilot Data Provision Notice (DPN), requires GP practices to pseudonymise their data held within the GPSS Systems and provide the Service with access to run approved queries against these pseudonymised patient records.

The Service is used to provide access to these stores of pseudonymised personal data to support Approved Projects related to the purposes outlined above. This Service provides users with the ability to remotely analyse and gain insight from linkable GP patient data within the GPSS boundary for approved research, clinical audit, service evaluation, health surveillance and health and social care policy, planning and commissioning purposes and public health purposes.

The Service will also be used to link pseudonymised record level data from other healthcare datasets obtained by NHS England, with the pseudonymised GP patient data held within the GPSS systems to support the **Purposes** (See Appendix 1, with additional technical details on the following website: Data Sources - OpenSAFELY documentation for full list of external datasets). All Approved Users will have data access agreements supported by NHS England.

The purpose is to provide a secure analytics service for Approved Users to access pseudonymised GP and other linked patient data for

- clinical audit[7];
- service evaluation[8];
- health surveillance[9];
- research[10];
- evaluation of the Service; and
- health and social care policy, planning and commissioning purposes and public health purposes, where agreed on a project specific basis by or on behalf of:
  - the Deputy Director of Data Policy on behalf of the Department of Health and Social Care, and
  - the Chief Data & Analytics Officer on behalf of NHS England, and
  - a nominated representative of each of the Royal College of General Practitioners and the British Medical Association on behalf of the Joint GP IT Committee.

NHS England and Approved Users need insights from this vital data to analyse and link data for the Purposes above, which will help improve the quality of people's lives and help save lives.

The Purposes for which this data may be analysed and used include:

- understanding risks to health, trends in diseases and such risks, and controlling and preventing the spread of diseases and such risks
- identifying and understanding information about patients or potential patients with, or at risk of disease
- understanding information about patient access to health services as a direct or indirect result of illness, and the availability and capacity of those services
- health services research on changes in clinical activity and population health and impact, as well as consequences of this
- informing the development and tracking the implementation of new diagnostics and treatments.

1. What is intended to be done with the personal data collected / used / processed / stored during this project?

Approved Projects will only be approved by NHS England for analysis for the Purposes, using the OpenSAFELY platform as a trusted research environment inside the GPSS.

The OpenSAFELY software performs the processing and provides results for Approved Projects.

The workflow for the Approved Project approvals process can be found here.

Personal data (and data relating to the deceased if they were registered with a GP practice on 1st January 2009) will be made available in pseudonymised form within the GPSS.

Both GPSS will produce a file that identifies their currently registered patients that do not wish their data from being shared outside the GP practice for planning and research

[7] As defined in the https://www.hra-decisiontools.org.uk/research/docs/DefiningResearchTable_Oct2022.pdf

[8] As defined in the https://www.hra-decisiontools.org.uk/research/docs/DefiningResearchTable_Oct2022.pdf

[9] As defined in the https://www.hra-decisiontools.org.uk/research/docs/DefiningResearchTable_Oct2022.pdf

[10] As defined in paragraph 3.1 pf the UK Policy Framework for Health and Social Care Research - Health Research Authority

purposes, known as a Type 1 Out Out (T1OO). The data in the file is pseudonymised and made available at Level 1, to ensure these patients are excluded from further data processing.

Approved Users will not have access to any of the pseudonymised GP or NHS England data, and most will only be able to view published anonymous, aggregated results after disclosure controls have been applied (the exceptions to "most" are a subset of approved users per project, who can view aggregated (and / or pseudonymised) results, inside the secure GPSS's environment, prior to the application of disclosure controls).

Trained output-checkers (employed by NHS England or through service contracts) will also be able to access the de-identified aggregated outputs to review so that disclosure controls that have been adequately applied by Approved Users; once cleared, these aggregated outputs are now considered anonymous and are released from the OpenSAFELY platform (i.e., to share with collaborators for discussion, or to incorporate as part of a manuscript or report).

Output checking is the final guarantee that outputs are aggregate and anonymous and cannot be used to re-identify patients.

Such Outputs are published to the "Job Server" dashboard outside the secure environment (https://jobs.opensafely.org/), from where they can be shared more widely with other project members and collaborators who do not have secure environment access.

Once approved, Approved Users can (and are encouraged to) self-publish their output-checked results (for example, as links to accepted manuscripts) for public consumption, within a peer-project dashboard that shows results alongside analytic code and audit logging.

2. How will the personal data be collected (i.e., will it be obtained from the individuals themselves or via a third party)?

Data will be collected from all GP Practices in England that use the GPSS that fall within the scope of the associated DPN. The GP practices use GP System Suppliers to store the patients' GP records which contain data that GPs have already obtained from patients and other third parties, including other healthcare professionals, for the purposes of providing healthcare services to patients.

Further data sources where NHS England is the owner and Data Controller will be provided and placed in the Level 2 environment (see data flow diagram) in Optum (formerly EMIS) and TPP alongside the GP data to allow Queries to run against both data sources. These data sources are pseudonymised to protect the identity of the patients but cannot be anonymised at this stage because they need to be linked using the pseudonym.

V1.1 Addition of Improving Access to Psychological Therapies to the list of permitted NHS England controlled datasets

- IAPT (Talking Therapies) is a fully onboarded product which is already used for research purposes
- The data will be provisioned along with the other NHSE data requirements via UDAL using recognised and agreed processes.
- There is an exit strategy in place if required
- The Cyber Security team have approved the security elements of the Processing and appropriate security documentation is in place. There are no new (since v1.0) cyber issues with OpenSAFELY. The organisation meet DSPT and Cyber

Essentials plus and were subject to a cyber assessment in 2025 with no concerns noted.

- The data only flows to GP System Suppliers (TPP and EMIS) (DPIA Section 4)
- There are NO new data engineering requirements
- The IAPT IAO has agreed the provision of IAPT in OpenSAFELY
- The data will only be made available under the NHS OpenSAFELY Data Analytics Service Pilot Directions 2025.

3. What will the intended results be, i.e., likely results for a GP Practice, impact (positive and negative, as applicable) on individuals concerned or (where applicable) other parties involved?

The data held within the GPSS environment are made available to the service to query, providing NHS England and Approved Users with insights from this data for the purpose outlined above**;** while also significantly reducing the burden of data requests on General Practice, enabling General Practitioners to focus on delivering health care and support to patients.

Consequently, users, and those who carry out research, clinical service evaluation, clinical audit, health surveillance, health and social care policy, planning, commissioning purposes and public health purposes will have more timely access to the data they require, assured through a transparent and robust governance process, providing scrutiny and transparency on the use of data about patients.

4. **What will be the benefits to the individuals concerned or (where applicable) other parties involved (including GP Practices) and to society?**

NHS England and Approved Users need insight from this vital data to analyse and link data for research, clinical service evaluation, clinical audit, health surveillance, health and social care policy, planning, commissioning purposes and public health purposes which will help save lives.

Supporting such research via the Service will reduce the burden on General Practice, the wider NHS (including hospital, community and mental health services), as well as provide increasing support to social care services; all at a time when demand on resources is high. We anticipate the insights will support General Practice and the wider health and care system to improve the quality and safety of care and support to patients better.

6. Description of the Processing

Nature and scope of the processing:

The Service is designed to keep patient data confidential: Approved Users write their analysis code (Query) away from the patient data; the Service automates the Query being run against the data sets. Only Aggregated Outputs and Log files are shared with users in the GP System Suppliers secure environment.

All actions performed within the Service are published in real-time at https://jobs.opensafely.org/. No record level patient data is shared outside of the GP system environment; disclosive checks ensure all Aggregated Outputs leaving the platform are anonymous.

A pseudonymised and further de-identified copy of coded GP patient data from the GP System Suppliers (see data flow diagram) is held in the Level 1 environment (one for Optum (formerly EMIS) and one for TPP). This copy of the GP data contains the coded and structured data of all patients in England registered at TPP or Optum practices.

All GP patient data is pseudonymised at source by the relevant supplier before being moved into the Level 1 environment. Further data sources are then ingested into each GP System Supplier's secure environment so that linkage can occur. The OpenSAFELY software can then perform queries on the data sets, following the Type 1 Opt-Out solution being implemented. The Services run Queries to generate final Aggregated Outputs (with disclosure controls applied) before any data leaves the GPSS' environment.

A small number of platform developers from the Bennett Institute (working under contract between the University of Oxford and NHS England) need more direct access to query the pseudonymised data (both GP – controlled and NHS England controlled data) inside the secure environment for systems integration and platform development, testing and maintenance purposes. This access is necessary to ensure the system is operating correctly for users, and controls are built in to ensure the required access is minimised, transparent and secure.

All developer access to the platform is logged and on a secure connection; the pseudonymised data developers can query is read-only; their queries are fully logged; there is no facility to enable the extraction of patient or record-level data outside of the environment (only Aggregated Outputs can be released). These activities, and the required access levels, are only undertaken by the Bennett Institute core platform development team who have role-specific security controls to reflect their increased access to data. Further information on the activities covered under these roles and the governing policies are available here: https://docs.opensafely.org/developer-access-policy/#policy-for-opensafely-access-by-platform-developers

Context of the processing:

Outputs will inform the health and care system on the provision of healthcare service and provide research insights to support wider health functions in England. Identifiable data is not extracted from the Service, rather anonymous and / or aggregated results are produced from the processing and provided to users.

7. Describe the legal basis for the processing (collection, analysis, or disclosure) of personal data?

1. NHSE's collection & analysis of personal data

i. Statutory Authority

NHS England is directed by the Secretary of State for Health and Social Care under section 254 of the Health and Social Care Act 2012 (**the 2012 Act**) to establish and operate a system for the collection and analysis of the information specified for this service. The Direction is titled The NHS OpenSAFELY Data Analytics Service Pilot Directions 2025 (**the Directions**) and is published on NHS England website.

These Directions are given in exercise of the powers conferred by sections 254(1), and 304(9), (10) and (12) of the 2012 Act[11] and sections 13ZC and 272(7) and 272(8) of the National Health Service Act 2006[12] (**the 2006 Act**).

The service enabled by these Directions is also operated by NHS England in the exercise of its powers under sections 1 and 2 and Chapter A1 of Part 2 of the 2006 Act and section 270 of the 2012 Act.

ii. Common law duty of confidentiality

Directed activity

Where we are directed to collect personal data, that legal obligation also acts as our legal basis for complying with the common law duty of confidentiality.

iii. UK GDPR lawful basis

Directed activity

UK GDPR Article 6(1)(c) - processing is necessary for compliance with a legal obligation to which the controller is subject (i.e. the Directions).

UK GDPR Article 9(2)(g) - processing is necessary for reasons of substantial public interest, on the basis of domestic law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject, by virtue of compliance with a direction supplemented by Data Protection Act 2018 (DPA 2018) Schedule 1, Part 2, paragraph 6: Statutory etc and government purposes.

2. NHSE's disclosure

i. Statutory legal basis

Our legal basis for disclosing personal data is section s.261 of the 2012 Act which allows us to share data as appropriate in order to comply with a direction under section 254 of the 2012 Act. Whether or not the sharing will be appropriate under s.261 will be assessed on a case by case basis depending upon the approved use case and nature of organisation applying for the data.

ii. Common law duty of confidentiality

The common law duty of confidentiality does not apply to the sharing of data as the data being shared is anonymous and no longer confidential.

iii. UK GDPR lawful basis

The UK GDPR does not apply to the sharing of data as the data being shared is anonymous.

3. GP's sharing of the Required Output

i. Statutory legal basis

[11] 2012 c.7 (the 2012 Act). Relevant amendments were made by the Health and Social Care Information Centre (Transfer of Functions, Abolition and Transitional Provisions) Regulations 2023 (S.I. 2023/98).

[12] 2006 c.41. Section 13ZC was inserted by section 45 of the Health and Care Act 2022 (c.31).

Section 259 of the 2012 Act gives NHS England the power to require and request the provision of information from other organisations. NHS England requires the necessary outputs from OpenSAFELY, on behalf of the GP practices, under s259(1)(a) of the 2012 Act.

As a result of s.259, GP's are under a legal obligation to provide the data in the form and manner described in this DPIA. In line with section 259(5) of the Act, all organisations in scope, in England, must comply with the requirement and provide information to NHS England in the form, manner and period specified in the associated NHS OpenSAFELY Data Analytics Service Pilot Data Provision Notice (DPN).

The above DPN is issued in accordance with the procedure published as part of an NHS England duty under section 259(8) of the 2012 Act.

ii. Common law duty of confidence

Where GPs are required to share the personal data with NHS England under s.259 of the 2012 Act, they can rely on that legal obligation to comply with the common law duty of confidentiality.

iii. UK GDPR lawful basis

The lawful basis under Article 6(1)(c) of the UK GDPR allows the GPs to share personal data as necessary under a legal obligation. The legal obligation is the requirement to share data under s259 of the 2012 Act.

The lawful basis for sharing special category data is UK GDPR Article 9(2)(g) which applies when processing is necessary for reasons of substantial public interest, on the basis of domestic law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject. The substantial public interest is compliance with a data provision notice issued to the GPs under s259(1)(a) of the 2012 Act by NHSE.

Under Article 9(2)(g) of UK GDPR, we need to identify the relevant provision under the DPA 2018. The relevant provision is Schedule 1, Part 2, paragraph 6 of the DPA 2018 (Statutory etc and government purposes) and Schedule 1, Part 2, paragraph 1 (the processing is necessary for healthcare purposes).

Criminal Convictions Data

In some circumstances, data from GP systems may relate to a criminal conviction which is processed on the platform. This data would likely be limited to the location of a patient at a particular time (e.g. if serving a prison sentence or being treated in a particular health and justice setting).

Where data is considered to be criminal convictions data, we rely on the same Article 6 and Article 9 lawful basis set out above as well as complying with Article 10 of the UK GDPR.

4. Restriction on publication

Where there is information collected under a Direction there is a duty to publish, unless it falls within one of the exclusions. By virtue of s260(2)(d) of the 2012 Act, NHS England is

directed by the Secretary of State not to publish data it obtains by virtue of the directions and therefore excluded.

8. Demonstrate the fairness of the processing

Lawfulness of processing: The service will only process data for the lawful basis specified above.

The fairness principle for processing is met as :

The purpose is defined within the Direction to provide a secure research service that utilises GP data in a manner that does not identify individuals and that any outputs are aggregated and / or anonymous.

The NHS England processing is transparent: Citizens will be informed of NHS England's processing through a transparency notice and the publication of the Directions Letters and associated Requirement Specification containing specific detail of the how the service will operate and the data that will be processed. The GP Practices will in addition, have transparency information for their patients regarding the service, including frequently asked questions.

To maintain the integrity and confidentiality of the data being processed: A combination of technical and process controls are in place. Including that the Service only uses pseudonymised data and only anonymous and aggregate data are released from the GPSS environment. These controls are detailed below in the section titled "What technical and organisational controls for information security have been put in place?"

The processing only allows the required data to be processed to support the service though data minimisation techniques being implemented: Any of the data items that are supported to be accessed via the OpenSAFELY technology will be subject to robust governance controls that will only allow the processing where it is necessary for the project. The data processed is only the data necessary to enable users to understand different risk profiles for people with different medical histories and demographic data, in relation to their health outcomes.

The data will not leave the GPSS environment and does not leave the boundary until its anonymous and or aggregated.

Patients can choose to opt out of their GP data being sharing outside their GP practice and therefore the data will not be available to the Service.

9. What steps have you taken to ensure individuals are informed about the ways in which their personal data is being used?

New transparency notices have been developed to communicate the changes to patients and issued to all GP practices with the new DPN. This content will be displayed by GP practices for their patients. This aligns with the NHS England focus on open data principles and the details contained in the General Transparency Notice available on the NHS England website.

Transparency information will be provided to GPs for their patients, and information about the OpenSAFELY data analytics data collection and processing is available on NHS England website https://digital.nhs.uk/about-nhs-digital/corporate-information-and-documents/directions-and-data-provision-notices/secretary-of-state-directions/nhs-opensafely-data-analytics-service-pilot-directions-2025/nhs-opensafely-data-analytics-service-pilot---privacy-notice?key=VRIOk17TG47NdtY0YiICSYwPCMsYYxjKH41TOMXOPuBzHnioLmx21b4BuUaP2xbR

NHS England will publish the Direction Letter, associated Requirement Specification and DPN on its website, once formally issued by the Secretary of State via the Department of Health and Social Care. The existing NHS England privacy notice for the OpenSAFELY Service will be updated in alignment with the changes described in this document.

Noting that the current data controllership position is complex, and there is potential for misunderstanding, transparency materials have been reviewed by NHS England, RCGP and BMA, and where necessary updated, to ensure the controllership arrangements are clear and easy to understand. In addition, consultation has occurred with relevant stakeholders as part of this pilot with opportunities to provide feedback.

NHS England has previously communicated to the public on the use of OpenSAFELY in the following two articles posted on the NHS England website:

- OpenSAFELY: secure access to data to deepen our understanding of COVID-19 - Improving care through research and innovation - NHS Transformation Directorate (england.nhs.uk)

- Through the NHS England Information Governance Portal – Questions on OpenSAFELY - IG frequently asked questions (FAQs) - Information governance - NHS Transformation Directorate (england.nhs.uk)

Separately, the OpenSAFELY website describes the software and its operation along with hosting a publicly available list of data sources: Data Sources - OpenSAFELY documentation. All research projects using OpenSAFELY are made public here: https://www.opensafely.org/approved-projects/.

10. Is it necessary to collect and process all data items?

Data Categories [*Information relating to the individual's*]	**YES/ NO**	**Justify** *[there must be justification for processing the data items. Consider which items you could remove, without compromising the purpose for processing]* ***FOR CLARITY: NOT ALL DATA ITEMS ARE USED IN EVERY CASE FOR EVERY APPROVED PROJECT.*** ***Any of the data items that are supported to be accessed via the OpenSAFELY technology are subject to robust governance controls, including the Project approval process, that will only allow the processing of specific data items, where it is necessary for the project***
Personal Data		
Name	NO	
Address	NO	
Postcode	YES	Agreed as part of project governance controls, on a project by project basis, as required **partial** postcodes are processed. It is necessary to identify risk factors, treatment options that might improve outcomes and approaches to treatment, and prevention of disease. E.g. areas of geographical depravation.
DOB	YES	Agreed as part of project governance controls, on a project by project basis, as required **partia**l DOB in format MM-YYYY. It is necessary to identify risk factors, treatment options that might improve outcomes and approaches to treatment, and prevention of disease. This variable informs calculation of age (see below).
Age	YES	Agreed as part of project governance controls, on a project by project basis, as required It is necessary to identify risk factors, treatment options that might improve outcomes and approaches to treatment, and prevention of disease.
Sex	YES	Agreed as part of project governance controls, on a project by project basis, as required It is necessary to identify risk factors, treatment options that might improve outcomes and approaches to treatment, and prevention disease.
Marital Status	YES	Agreed as part of project governance controls, on a project by project basis, as required It is necessary to identify risk factors, treatment options that might improve outcomes and approaches to treatment, and prevention of disease. Through the identification and processing of the data item necessary for approved projects subject to robust governance arrangements being implemented. Many ethical and valid studies exist in journals that have processed marital status, so the use of such a code is not contentious nor novel; and like many studies, researchers will discuss in their papers the limitations and strengths of their studies, which where necessary will include aspects related to marital status.
Gender	YES	Agreed as part of project governance controls, on a project by project basis, as required. It is necessary to identify risk factors, treatment options that might improve outcomes and approaches to treatment, and prevention of disease.

Data Categories [*Information relating to the individual's*]	**YES/ NO**	**Justify** *[there must be justification for processing the data items. Consider which items you could remove, without compromising the purpose for processing]* ***FOR CLARITY: NOT ALL DATA ITEMS ARE USED IN EVERY CASE FOR EVERY APPROVED PROJECT.*** ***Any of the data items that are supported to be accessed via the OpenSAFELY technology are subject to robust governance controls, including the Project approval process, that will only allow the processing of specific data items, where it is necessary for the project***
Living Habits	YES	Agreed as part of project governance controls, on a project by project basis, as required It is necessary to identify risk factors, treatment options that might improve outcomes and approaches to treatment, and prevention of disease. E.g. smoking, alcohol consumption, exercise habits
Professional Training / Awards / Education	NO	
Income / Financial / Tax situation / Financial affairs	NO	
Email Address	NO	
Physical Description	YES	Agreed as part of project governance controls, on a project by project basis, as required. It is necessary to identify risk factors, treatment options that might improve outcomes and approaches to treatment, and prevention of disease.
General Identifier e.g., NHS No	YES	Agreed as part of project governance controls, on a project by project basis, as required. NHS Numbers are pseudonymised to protect the identity of patients it is used to link datasets within the secure data environment.
Home Phone Number	NO	
Online Identifier e.g., IP Address/Event Logs	NO	
Website Cookies	NO	
Mobile Phone / Device No / IMEI No	NO	
Location Data (Travel / GPS / GSM Data)	NO	
Device MAC Address (Wireless Network Interface)	NO	
Banking information e.g., account number, sort code, card information	NO	
Special Category Data		
Physical / Mental Health or Condition	YES	Agreed as part of project governance controls, on a project by project basis, as required. It is necessary to identify risk factors, treatment options that might improve outcomes and approaches to treatment, and prevention of disease. Data is collected and processed except: a. SNOMED Refset for 'General Practice summary data sharing exclusion for gender related issues' 999004371000000109 b. SNOMED Refset for 'General Practice summary data sharing exclusion for assisted fertility' 999004351000000100 c. SNOMED Refset for 'General Practice summary data sharing

Data Categories [*Information relating to the individual's*]	**YES/ NO**	**Justify** *[there must be justification for processing the data items. Consider which items you could remove, without compromising the purpose for processing]* ***FOR CLARITY: NOT ALL DATA ITEMS ARE USED IN EVERY CASE FOR EVERY APPROVED PROJECT.*** ***Any of the data items that are supported to be accessed via the OpenSAFELY technology are subject to robust governance controls, including the Project approval process, that will only allow the processing of specific data items, where it is necessary for the project***
		exclusion for termination of pregnancy' 999004361000000107 d. All children of the SNOMED code 118199002 'Finding related to sexuality and sexual activity'
Sexual Life / Orientation	Partial	Agreed as part of project governance controls, on a project by project basis, as required. It is necessary to identify risk factors, treatment options that might improve outcomes and approaches to treatment, and prevention of the disease. [Placeholder, link to the website https://nhsengland.kahootz.com/t_c_home/grouphome
Religion or Other Beliefs	YES	Agreed as part of project governance controls, on a project by project basis, as required. It is necessary to identify risk factors, treatment options that might improve outcomes and approaches to treatment, and prevention of disease.
Trade Union membership	NO	
Racial / Ethnic Origin	YES	Agreed as part of project governance controls, on a project by project basis, as required. It is necessary to identify risk factors, treatment options that might improve outcomes and approaches to treatment, and prevention of disease.
Biometric Data (Fingerprints / Facial Recognition)	NO	
Genetic Data	NO	We believe the answer is "no": we may have access to genetic diagnostic codes, for example stating that someone has a disease with a genetic component (e.g., "Down's Syndrome") but we do not have access to individuals' complex genome or actual gene sequencing information as these are not typically stored as structured or coded data in GP records.
Criminal convictions / alleged offences / outcomes / proceedings / sentences	YES	Agreed as part of project governance controls, on a project by project basis, as required. As ODS/ SMOMED code, not free text, where a GP has (infrequently) chosen to capture such coded information relating to this as part of the health record.

Publicly available list of data sources: Data Sources - OpenSAFELY documentation

11. Describe if personal datasets are to be matched, combined, or linked with other datasets (internally or for external customers)

The GP data will be matched to NHS England controlled datasets as defined in Appendix 1. The NHS England controlled datasets and GP data can be matched through the

common pseudonymisation process; it should be noted that any Type 1 Opt-Outs are maintained within the GP dataset.

What follows is an explanation of how the process works

NHS England uses two methods for matching the pseudonym.

1. **Matching by external data provider**: Optum (formerly EMIS) or TPP sends a file of all their patient pseudonyms to the external data providers; only patient records with a match in the external databases are transferred back into the Level 2 environments in Optum and TPP. This is the preferred mechanism as it minimises unnecessary data flow.
2. **Matching by the GPSS:** On some occasions (due to technical constraints of the external database provider or when the external provider's population is much smaller, relative to that of the GP population), the pseudonym matching occurs inside the Level 2 environment in Optum and TPP. This involves a two-step process:
    a. The external data provider makes available one file of their full pseudonym list to the GPSS which in turn establish a list of matches of the pseudonyms they have. Each GPSS sends back a list of these matched pseudonyms to the external data provider in a file.
    b. The external data provider then prepares separate files for each of the data sets according to the matched list they were provided with. The two separate files produced by the external data provider, one for EMIS and one for TPP, contain the additional requested and approved data for only the matched pseudonyms. This data is then made available to be added to the Level 2 environment in Optum and TPP.

For record linkage, the pseudonyms will be produced by the GPSS who will hold the keys. This pseudonym (with additional data items used where necessary) provides significant confidence in matching accuracy, ensuring that the right record is linked to the right primary care record. Research models will be automatically executed as required for analysis. GP patient data held by the GPSS is incrementally updated to reflect changes made by clinicians in England GP practices.

The patient data processed for this purpose includes:

- Demographic information (age, sex, area of residence, ethnicity),
- Clinical information pertaining to wider health conditions, medication, allergies, physiological (e.g., BMI), prior blood tests and other investigation results, and other recent medical history (e.g., smoking status).

The patient population made available within the OpenSAFELY Databases in the GPSS environments for analysis is defined as all patients EXCEPT:

- Patients who have no period of registration (in a TPP or EMIS practice) after 1 January 2009 or who died before 1 January 2009 (year of Swine Flu Pandemic)

This means that for each system supplier if there has been any period of registration after 1 January 2009 then all the patient's data from the last practice with that system supplier is available to the Service for analysis, irrespective of the patient's current registration status.

Patients' Type 1 Opt-Outs will be respected (see Type 1 Opt-Out heading in Section 4).

Patient data made available to NHS England via the data stores in the GPSS environments is pseudonymised and further de-identified at source.

For this patient population, all coded patient data will be made available for processing, except the following codes in reference sets - SNOMED CT - NHS England Digital[13]

A summary of the specific data sets that are matched, combined, or linked are described in Appendix 1 - NHS England Controlled Datasets. All linking happens within The NHS OpenSAFELY Data Analytics Pilot Service.

The transfer mechanism varies on the data provider and outlined in Appendix 1

Data Sharing Agreements are in place to support the current process of linking GP data and NHS England controlled datasets data flowing to both data stores (in Optum and TPP). The data store in TPP contains GP data from TPP, linked to (for example) ONS Deaths; and the data store in EMIS contains GP data from Optum, linked to the same datasets. These two data stores are not combined together - they are kept separate and continue to reside in the separate GPSS environments, each having a separate implementation of the OpenSAFELY platform.

The process of ingesting pseudonymised NHS England data may be altered in the future according to the wishes of NHS England. Any changes will result in an update to this DPIA.

All externally linked datasets currently flow into Level 2 environments in TPP and Optum.

12. Describe if the personal data is to be shared with other organisations and the arrangements you have in place

Personal data is not shared with other organisations.

Whilst the Service processes pseudonymised data, all outputs released from the Service are in aggregated and / or anonymous form, with disclosure controls applied in accordance to the Code of Practice for Statistics.

The Approved User will submit a query (that was prepared on randomly generated dummy data) which is then run against the pseudonymised personal data held securely within TPP and Optum (formerly EMIS). This query returns an Aggregated Output and any logs (if produced - see Section 2 Background), which are only viewed inside the secure environment of the GP System Suppliers (TPP and Optum). Before any Aggregated Outputs are released outside the GP System Supplier environment (for sharing amongst the wider research team or collaborators, or for publication) Approved Users apply statistical disclosure controls and the outputs are reviewed and cleared by trained output-checkers; such outputs are now considered anonymous. Therefore no “personal” data is disseminated by design outside of the GP System Supplier environments. The Approved Users only access Aggregated Outputs and Log files inside the GP System Supplier environment and their access is logged.

- [13] SNOMED Refset for ‘General Practice summary data sharing exclusion for gender related issues’ 999004371000000109
- SNOMED Refset for ‘General Practice summary data sharing exclusion for assisted fertility’ 999004351000000100
- SNOMED Refset for ‘General Practice summary data sharing exclusion for termination of pregnancy’ 999004361000000107
- All children codes of the SNOMED code 118199002 ‘Finding related to sexuality and sexual activity’.

The above-restricted codes reflect current guidance regarding restrictions around the sharing of certain patient information and can be updated to align with new NHS England and professional guidance. All children codes of the SNOMED code 118199002 ‘Finding related to sexuality and sexual activity’.

13. How long will the personal data be retained?

All data will be held according to UK GDPR Regulations and DPA 2018. The retention period is in line with the Records Management NHS Code of Practice. Furthermore, the NHS England Records and Document Management Policy and Corporate Retention and Disposal Framework Implementation Process will be adhered to.

As the NHS England data stores are held with the data processors TPP and Optum (formerly EMIS), the data will be held for verification of findings and audit purposes and can be deleted once outside the retention periods, however the data cannot be transferred to other NHS England systems (as per operating requirements set out by the Secretary of State). The Intermediate Outputs may also be retained for verification of analyses and for audit purposes.

If a currently registered patient registers or de-registers a Type 1 Opt-Out this would take effect at the time of the next pseudonymised database build by the GPSS, usually within one week and be reflected in the results queries. Where queries have already been run prior to the Type 1 Opt-Out being applied and Intermediate Outputs have been generated (e.g., Level 3 onwards), these data sets would continue to hold the Type 1 data to ensure the integrity of the research study that ran the query.

14. Where you are collecting personal data from the individual, describe how you will ensure it is accurate and if necessary, kept up to date

The NHS OpenSAFELY Data Analytics Pilot Service does not collect personal data from individuals; the personal data which is processed through the service has been obtained or generated by GP Practices (for GP data) or as been obtained from other data controller organisations in the case of NHS England data.

15. How are individuals made aware of their rights and what processes do you have in place to manage such requests?

Individuals (data subjects) have the following rights under GDPR:

- The right to be informed – Fair Processing information and Transparency Notice for NHS England and GPs have been developed by NHS England and made available to GP Practices to support their obligations under UK GDPR. These are aligned accordingly to reflect the description in the Direction and associated Requirement Specification. A specific privacy notice is available on the NHS England website. In

addition, all research projects using OpenSAFELY software are made public here: https://www.opensafely.org/approved-projects/

- The right of access (NHS England) - An explanation about how an individual can request a copy of information that NHS England holds is published at: https://digital.nhs.uk/article/6851/How-to-make-a-subject-access-request.

  Due to the pseudonymisation of the NHS England data held within the GPSS environments, it would require a re-identification of the records to provide details of the data used in the Service. Consequently, NHS England can provide information of the source data that NHS England holds related to an individual, but will not be able to identify if a specific individuals data has been made available for specific Approved Projects.

- The right of access (GP Practices) - Any patient can make a subject access request to see part or the whole of their medical records from the GP Practice. Information about how to make these requests should be available on GP Practice websites. To minimise the burden on GPs at this time patients are encouraged to register and use NHS App services which include access to medical records.

- The right of access (other sources) - For data originating from non-NHS England sources, patients can make their requests directly with the source providers (as noted on the following website - Data Sources - OpenSAFELY documentation).

- The right to rectification - The right for individuals to have inaccurate personal data rectified, or completed if it is incomplete. NHS England via the Service cannot uphold this right, as the Service does not re-identify patients; however, individuals can make a request to either the GP Practices or the source data provider to rectify any errors.

- The right to erasure - the right of erasure does not apply as the legal basis of processing under UKGDPR is not consent or legitimate interest.

- The right to restrict processing - Where an individual contests the accuracy of their personal data NHS England will consider the request..

- The right to data portability- is not applicable to this processing because under article 20 (3) the processing is being carried out in the exercise of official authority vested in the controller under Article 6(1)(c) legal obligation under the NHS OpenSAFELY Data Analytics Service Pilot Directions 2025 or under Article 6(1)(e) public task.

- The right to object – is not applicable to this processing as data is being processed under legal obligation
- Rights in relation to automated decision making - no automated decision making takes place as part of this processing.

16. What technical and organisational controls for "information security" have been put in place?

The governance and service model includes: Application stage – applications are assessed to ensure that:

- The application is for research, clinical service evaluation, clinical audit, health surveillance and health and social care policy, planning and commissioning purposes and public health purposes.
    - Applications for a health and social care policy, planning and commissioning purposes and public health purposes, where agreed on a project specific basis by or on behalf of:
        - the Department of Health and Social Care,
        - NHS England, and
        - a nominated representative of each of the Royal College of General Practitioners and the British Medical Association on behalf of the Joint GP IT Committee
- The data necessary to support the purpose of the application is available in the system.
- The applicant has submitted a completed application form, along with any relevant supporting documentation
- If the application is for research, that a favourable opinion is provided by an NHS Research Ethics Committee (REC) where required;
- If the application is for clinical audit, service evaluation and health surveillance purposes, that the purpose has had review from a local or institutional ethics committee (to ensure the purpose it is appropriately categorised), or the HRA decision tool is used[14] where there is no local or institutional ethics committee to provide review.

  Further detail surrounding the OpenSAFELY application process can be found here: https://jobs.opensafely.org/apply/
- Technical controls- Data is pseudonymised and further de-identified at source; linkage of datasets is managed inside by the GPSS systems. The requirement to exclude patients' data for analysis and the Purposes through Type 1 Opt-Outs is applied through the OpenSAFELY software. The OpenSAFELY software requires access to a pseudonymised list of patients who have registered a Type 1 Opt-Out. Their opt-out is upheld (i.e. their data is not made available to any Approved Projects) when a Query is run.

- No system administrators or platform developers have access to the pseudonymisation key. No event level or patient level data leaves the service (i.e., the secure network boundaries of the GPSS' environments). Users must specify up front the code they are using to analyse the patient data: explicitly writing studies as "analysis code" means that it is possible for any interested party to check exactly how patient data was processed, and to assess if such processing is in line with the approved project purpose; the platform principles also require that all analysis code is made public. It is accepted that some code may remain private while an analysis is in development. However, all code is published when the results of the analysis

[14] https://www.hra-decisiontools.org.uk/research/

are shared (or, for non-complete projects, as soon as possible, usually at the point of their cessation, and no later than 12 months after any code has been executed against the patient data). System administrators can control when analysis code is made public to maintain our transparency principles. Approved Users can only access their study Aggregated Outputs and any Logs (if produced) over a secure encrypted connection. Access to the Aggregated Outputs is via a secure encrypted connection, unique to each user, with all access audited. In addition, all researcher actions in the Service are logged in public, in real-time and all Queries are logged and published (https://jobs.opensafely.org/).

- People controls – All Approved Users must pass safe researcher training (such as that provided by the ONS or UK Data Service) to have access to the secure environment (Level 4) hosting the Aggregated Outputs. All Approved users who write study code or access the Level 4 environment, and all output checkers, must sign a Data Access Agreement approved by NHS England. All Approved new users have a 'co-pilot' who is an expert user providing appropriate training and support to ensure safe practice is followed and plausible results are generated.

- Output checking – Approved Users apply disclosure controls to their study aggregate outputs they request for release from the secure environment. All aggregate outputs are then independently checked to ensure that they are non-disclosive and safe to release by output checkers who have been suitably trained.

- Transparency – all code run using the system is published online (https://jobs.opensafely.org/), and Approved Users are asked to share links to the papers, reports, blogs and other material (such as presentations) that have been approved for publication on their public facing project page (project pages can be found by following links from these respective organisations,) https://jobs.opensafely.org/organisations/. An example project page showing a link to a paper is found here: https://jobs.opensafely.org/comparative-vaccine-effectiveness/). A list of key papers published in peer-reviewed journals is listed here: https://www.opensafely.org/research. All accepted applications have their study purpose, responsible organisation and study lead information published online.

The data stores in Optum (formerly EMIS) is based on AWS Cloud data centres, which are certified for compliance with ISO/IEC 27001:2013 (IT - security techniques - information security management systems), 27017:2015 (IT - security techniques – code of practice for information security controls based on ISO/IEC 27002 for cloud services), 27018:2019 (IT - security techniques - code of practice for protection of personally identifiable information (PII) in public clouds acting as PII processors) and ISO/IEC 9001:2015 (quality management systems). Optum has obtained a Standards exceeded DSPT submission for 2025. Certification details are available at https://aws.amazon.com/compliance/iso-certified/. AWS has obtained a 'Standards Exceeded' Data Security and Protection Toolkit (DSPT) submission (quality management systems). Certification details are available at https://aws.amazon.com/compliance/iso-certified/.

Data in transit will be protected by standard security policies available for AWS Elastic load balancers and application load balancers (https://docs.aws.amazon.com/elasticloadbalancing/latest/network/create-tls-listener.html)

For the TPP data store the TPP infrastructure involves a Tier 3 data centre accredited to NHS England standards for centrally hosted clinical systems and a secure office environment. TPP is accredited to the ISO 27001 standard and has obtained Standards Exceeded DSPT submission in 2025. The same governance, security and audit protocols that apply to the production TPP SystmOne environment will extend to cover this data store. This includes the security and audit arrangements required for the Overarching Digital Services for Integration Standards and the Health & Justice Information Services programme (HJIS), for example.

Pseudonymisation, de-identification and minimisation at source (Process)

No free text information or directly identifiable patient information is ever transferred to the Service; redundant data is pruned (for example, the datasets available to the NHS OpenSAFELY Data Analytics Pilot Service do not contain the full data schema for SUS data, but a subset); and data about patient geographic locations is always rounded up to the least disclosive level that is still analytically useful (typically, Middle-Layer Super Output Area (MSOA) level for patients, and Integrated Care System level or Local Authority code for organisations). The level of data available is all publicly documented. Identifiable patient information within the GPSS environments (and other data provider organisations) is further de-identified according to the processes outlined in this DPIA (above). In all cases, a pseudonym is created with the patient's NHS number in combination with a salt using the SHA2 512 cryptographic hash technique. The SALT is shared by the GPSS with the external data provider: it is emailed as an encrypted file to a specific individual in the external data provider organisation post approval from NHS England. A member of either GPSS then calls the individual and provides the decryption key to unencrypt the salt.

There is no directly identifiable data being processed under the scope of this DPIA. It is acknowledged that the sources of the data being provided (e.g., GP Practices) have access to the identifying data, but that is outside of the scope of this DPIA.

The OpenSAFELY analytics tool operates a tiered security model, described here: https://docs.opensafely.org/security-levels/.

Identifiable data is accessible only in Level 1: this is not available to any Approved users or developers. Level 1 is the data held within the direct care servers of the GPSS, who operate as data processors for GP practices, or the external source data provider environments. In addition, Level 1 is where the pseudonymised and further de-identified GP data is created and stored, which remains under the control of GP practices. The Service automates the running of code (the Queries) against pseudonymised GP data (Level 1) and pseudonymised NHS England data (Level 2), to generate intermediate pseudonymised datasets (Intermediate Outputs) in Level 3, and then final anonymous aggregated outputs (the Aggregated Outputs) in Level 4 (See Data Flow Diagram).

Pseudonymised and aggregate data is accessed via a secure encrypted connection provided by the system supplier, within their main warehousing infrastructure. Only aggregated data (after disclosure controls have been applied, as part of the Five Safes Framework) leave the secure environments within Optum and TPP.

17. In which country/territory will personal data be stored or processed?

The UK.

The geographical location of the data once transferred to the GPSS is within the UK jurisdiction in their secure environments.

All NHS England staff and contractors accessing the data store will be doing so from within the UK. The OpenSAFELY software platform was built by the Bennett Institute and operates from the UK.

The NHS England Data Sharing Agreements (for the OpenSAFELY data) restricts the use to defined territories. The application process includes assessing the recipient's legal basis for processing data within these territories. This is reviewed and approved by NHS England according to the UK GDPR, DPA 2018 and the NHS Records Management : Code of Practice.

18. Does the National Data Opt Out apply to the processing?

No. The Service through the OpenSAFELY Data Analytics Pilot Direction will be lawfully enabled to process pseudonymised data for the Purposes previously stated.

The National data opt-out (NDOO) operational policy guidance document[15], Chapter 7 details the policy considerations for specific organisations or purposes. Paragraph 7.10 relates to application of NDOO by NHS England (in relation to the powers it exercises which were formerly those of NHS Digital). NDOO is exempt when data is required by NHS England under section 259 of the 2012 Act, following a Direction to collect it.

The Service also does not process Confidential Patient Information, the GPSS's pseudonymise the data, under instruction of the GP in order to comply with a request from NHS England under section 259 of the 2012 Act, prior to a Query being run and the pseudonymised data being made available to NHS England in the form of Intermediate Outputs. Subject to output checking, only anonymous and aggregate data is shared with Approved Users of the Service. The NDOO does not apply to anonymous data.

Patients that have a Type 1 Opt-Out code in their records at their currently registered practice will be upheld and not have their data processed by the NHS OpenSAFELY Data Analytics Pilot Service. This is consistent with other NHS England GP data collections directed under section 254 of the 2012 Act, for planning and research purposes.

In certain limited circumstances and where ethics approvals support it, a project may wish to apply the NDOO as part of the Query they have developed, notwithstanding that the Service operates under an exemption to the National Data Opt Out Policy. NHS England will therefore work with GP System Suppliers to develop a technical solution to enable the application of the NDOO at project level at the request of an Approved User in these circumstances.

[15] National data opt-out operational policy guidance document - https://digital.nhs.uk/services/national-data-opt-out/operational-policy-guidance-document

19. Identify and assess risks

Please see FINAL DPIA Risk Assessment-V1.0 to review all risks

20. Actions required as identified in this DPIA

Action No	***Actions required*** *(Date and responsibility for completion)*	***Risk No impacted by action***	***Action owner*** *(Name and role)*	***Date completed***
1	*End to end cyber assurance review of the OpenSAFELY platform with a report delivered to OS Programme Board.*	*OS-DPIA-007*	*REDACTED* *Head of Programme Delivery*	*July 2025*
2	*Development and implementation of an OpenSAFELY service governance framework*	*OS-DPIA-008*	*REDACTED* *Head of Programme Delivery*	*September 2025*
3	*Review and ensure the OpenSAFELY Service operates in compliance with the NHS Records Management; Code of Practice and NHS England Records Management Policy*	*OS-DPIA-014*	*REDACTED* *Head of Programme Delivery*	*September 2025*

Action No	***Actions required*** *(Date and responsibility for completion)*	***Risk No impacted by action***	***Action owner*** *(Name and role)*	***Date completed***
4	*Data Processing Agreement with Optum to be developed and signed.*	*OS-DPIA-020*	*REDACTED* *Head of Programme Delivery*	*July 2025*
5	*Publication of the Direction Letter, associated Requirement Specification and DPN to the NHS England website*	*OS-DPIA-002*	*REDACTED* *Head of Data Governance & Assurance*	*June 2025*
6	*Publication of the NHS England DPIA and documentation developed to support GP practices (GP DPIA template, GP IG FAQ's)*	*OS-DPIA-002*	*REDACTED* *Head of Data Governance & Assurance*	*July 2025*
7	*NHSE to put in place appropriate governance procedures with OpenSAFELY for reviewing any project specific requests for NDOO to ensure that NDOOs are applied appropriately in line with the National Data Opt Out Policy at project level.*	*OS0-DPIA-022*	*Michael Chapman* *Director of Data Access and Partnerships*	*September 2025*

Action No	***Actions required*** *(Date and responsibility for completion)*	***Risk No impacted by action***	***Action owner*** *(Name and role)*	***Date completed***
8	*A programme plan to be developed and implemented for auditing and the undertaking of appropriate assurance of the OpenSAFELY service.*	*OS-DPIA-018*	*REDACTED* *Head of Programme Delivery*	*December 2025*

V.1.1 – TO NOTE: no additional risks identified with regards to the addition of Improving Access to Psychological Therapies to the list of permitted NHS England controlled datasets

21. Further Actions

- The completed DPIA should be submitted to the PTT Helpline Service (ighelplineservice@nhsdigital.nhs.uk) for review
- The IAO (Information Asset Owner) should keep the DPIA under review and ensure that it is updated if there are any changes (to the nature of the processing and/or system changes)
- A redacted version of the DPIA should be made available to the public

22. Signatories

The DPIA accurately reflects the processing and the residual risks have been approved by the Information Asset Owner:

REDACTED signature Director of Data Access and Partnerships 14/07/2025 V1.1 via email 15/01/2026 V1.2 via email 13/03/2026

Information Asset Owner (IAO) Signature and Date

FOR PRIVACY, TRANSPARENCY AND TRUST AND OFFICE OF THE DPO USE ONLY

23. Summary of high residual risks

Risk no.	High residual risk summary

Summary of DPO advice:

Data Protection Officer (DPO) **Signature and Date**

ICO consultation outcome:

Office of DPO **Signature and Date**

Next Steps:

- **DPO to inform stakeholders of ICO consultation outcome**
- **IAO along with DPO and SIRO to build action plan to align the processing to ICO's decision**

24. Appendices.

24.1. Appendix 1 - NHS England Controlled Datasets

Sender	*Content*	*Pseudonymised?*	*Mode*	*Security*	*Recipient*
A report that describes the dataset flows (name and when updated) for TPP can be found here: https://github.com/opensafely/database-notebooks/blob/master/notebooks/database-builds.ipynb Details for dataset flows with EMIS will be available shortly.					
NHSE	SUS data (APCS, ECDS, OPA)	YES	Encrypted electronic transmission, using industry standard encryption methods.	Encrypted file of de-identified data encrypted using industry standard methods.	TPP and EMIS
NHSE	IAPT	YES	Encrypted electronic transmission, using industry standard encryption methods.	Encrypted file of de-identified data encrypted using industry standard methods.	TPP and EMIS
NHSE	NHS Civil Registration Data in England & Wales	YES	Encrypted electronic transmission, using industry standard encryption methods.	Encrypted file of de-identified data encrypted using industry standard methods.	TPP and EMIS

24.2. Appendix 2 - **Example Intermediate Output tables:**

	Has Diabetes	HbA1c level	Date of death	Prescribed antivirals	Admissions
1234	1	45.2	30/02/2023	2	1
5678	1	40.1	-	0	1
2345	0	38.2	-	1	0

3456	0	40	04/08/2019	0	3
5667	1	68.3	-	0	2

In the above table the patient level data may include reference to a specific patient event depending on the nature of the study and the clinical events experienced by the patient. For example, in the table, patients 1234 and 5678 both had 1 event of an admission during the study time period; patient 2345 had one event of antiviral prescriptions; the date of the single clinical event “death” is given, as this is essential to the design of the statistical analysis. Note that in the creation of this dataset the total patient record has been curated and minimised down to the level of detail required for the specific study being executed.

In some circumstances, depending on the nature of the study, Approved Users may reshape their Intermediate Output tables into a different format because this allows the OpenSAFELY platform to run queries more efficiently. An example is below:

If we’re interested in covid tests, we might need to know the results for multiple tests. We can extract this into a “wide” one-row-per-patient table with one column for each date in our study period.

patient	20230310	20230311	20230312	20230313	20230314
1234	neg	-	-	-	-
2345	-	neg	neg	pos	-
3456	-	-	neg	-	-
4567	-	-	pos	-	-
5678	-	-	-	-	neg

This information might be more efficiently represented by extracting it into Level 3 as a “long” one-row-per-event table; importantly, this is the same information. Note that in the creation of this dataset the total patient record has been curated and minimised down to the level of detail required for the specific study being executed.

patient	val	date
3456	neg	2023-03-12
2345	neg	2023-03-12
1234	neg	2023-03-10
4567	pos	2023-03-12
2345	neg	2023-03-11
5678	neg	2023-03-14
2345	pos	2023-03-13

The list of pseudonymised datasets held within the Service can be found on the following website - https://docs.opensafely.org/data-sources/ [Placeholder – NHSE website]

More information about how the Service operates is available on the OpenSAFELY website here- https://www.opensafely.org/