

Document filename:	Annex A - General Practice Data for Planning and Research Requirements Specification		
Project / Programme	General Practice Data for Planning and Research	Project	
Document Reference	<insert>		
Project Manager	Tess Morley	Status	Final
Owner	Eva Simmonds	Version	1.0
Author	Tess Morley	Version issue date	01/04/2021

Annex A - General Practice Data for Planning and Research Requirements Specification

Document management

Revision History

Version	Date	Summary of Changes
0.1	23/10/2020	Initial draft
0.2	28/10/2020	Subsequent draft incorporating feedback from reviewers
0.3	02/11/2020	Minor changes
0.4	05/11/2020	Subsequent draft incorporating feedback from reviewers
0.5	07/12/2020	Changes following review by Executive Director of Privacy, Transparency and Ethics
0.51	10/12/2020	Snapshot draft to support NHSX review process
0.6	18/12/2020	Subsequent draft incorporating feedback from reviewers
0.7	04/02/2021	Subsequent draft incorporating feedback from Executive Director of Privacy, Transparency and Ethics
0.8	10/02/2021	Subsequent draft incorporating feedback from the British Medical Association, Royal College of General Practitioners and Executive Director of Privacy, Transparency and Ethics
0.9	16/02/2021	Final draft incorporating all feedback and minor changes.
1.0	01/04/2021	Final version for publication

Reviewers

This document must be reviewed by the following people:

Reviewer name	Title / Responsibility	Date	Version
Richard Birmingham	Senior IG Specialist Advisor - Data Protection and IG Compliance	28/10/2020	0.1
Peter Short	Clinical Lead for General Practice Data for Planning and Research	28/10/2020	0.1
Stephanie Tod	Communications Business Partner for General Practice Data for Planning and Research	28/10/2020	0.1
Carol Lodge	Business and Operational Delivery Manager, Primary Care Domain	28/10/2020	0.1
Stuart Cumming	Senior Project Manager, General Practice Data for Planning and Research	02/11/2020	0.2
Magi Nwolie	Principal IG Specialist and Associate Data Protection Officer, Data Protection and IG Compliance	03/11/2020	0.3
Peter Short	Clinical Lead for General Practice Data for Planning and Research	03/11/2020	0.3
Jackie Gray	Executive Director of Privacy, Transparency and Ethics	07/12/2020	0.4
Richard Birmingham	Senior IG Specialist Advisor - Data Protection and IG Compliance	08/12/2020	0.5

Peter Short	Clinical Lead for General Practice Data for Planning and Research	08/12/2020	0.5
Andrew Thorne-Marsh	Programme Manager, General Practice Data for Planning and Research	08/12/2020	0.5
Andrew Thorne-Marsh	Programme Manager, General Practice Data for Planning and Research	14/12/2020	0.51
Carol Lodge	Business and Operational Delivery Manager, Primary Care Domain	15/12/2020	0.51
Richard Birmingham	Senior IG Specialist Advisor - Data Protection and IG Compliance	15/12/2020	0.51
Peter Short	Clinical Lead for General Practice Data for Planning and Research	15/12/2020	0.51
Jackie Gray	Executive Director of Privacy, Transparency and Ethics	16/01/2021	0.7
Peter Short	Clinical Lead for General Practice Data for Planning and Research	16/02/2021	0.9
Andrew Thorne-Marsh	Programme Manager, General Practice Data for Planning and Research	16/02/2021	0.9
Eva Simmonds	Programme Head, General Practice Data for Planning and Research	23/03//2021	0.9
Jackie Gray	Executive Director of Privacy, Transparency and Ethics	23/03//2021	0.9

Approved by

This document must be approved by the following people:

Name	Signature	Title	Date	Version
Eva Simmonds		Programme Head	01/04/21	1.0

Glossary of Terms

Term / Abbreviation	What it stands for
BMA	British Medical Association
DAE	Data Access Environment
DARS	Data Access Request Service
DHSC	Department of Health and Social Care
DPS	Data Processing Services
GPES	General Practice Extraction Service
GDPR	General Data Protection Regulation

IGARD	Independent Group Advising on the Release of Data
JGPITC	Joint GP IT Committee - a sub-group of the RCGP formed of GPs with specialist interests in data and IT and policy leads and ethics representatives from the BMA
RCGP	Royal College of General Practitioners

Document Control:

The controlled copy of this document is maintained in the NHS Digital 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.

Purpose of document

This document sets out the requirements for the General Practice Data for Planning and Research Directions Specification and should be read alongside the:

- General Practice Data for Planning and Research Directions 2021 (the Directions) issued by the Department of Health and Social Care (DHSC) on behalf of the Secretary of State for Health and Social Care
- Annex B – General Practice Data for Planning and Research Data Set Specification
- Data Provision Notice
- Data Protection Impact Assessment
- Transparency Notice

Introduction

Background

NHS Digital currently fulfils its statutory duties to collect secondary use data from GP practices through the General Practice Extraction Service (GPES). In 2014, a National Audit Office review of GPES found that the service was inefficient and costly but fulfilled a necessary requirement and should be replaced and improved. A subsequent Public Accounts Committee hearing in Parliament upheld and supported these findings. GPES is inefficient, costly and capacity constrained because it is a “collect once, use once” method, with limited automation.

The policy demand and commission were for the replacement to support existing national extracts and reports, in addition to providing capability and capacity to meet rapidly increasing requirements for data to support local, regional and national planning, commissioning and research. It was agreed that this should be achieved in a way that builds and maintains professional, patient and public trust, and operates within the legal, policy and transparency framework of NHS Digital.

In September 2018, a Full Business Case was approved for the development of General Practice Data for Planning and Research to replace GPES for secondary use purposes. Full replacement of GPES capability will require the inclusion of data from GP medical records which have a Type 1 opt-out recorded on them. Type 1 opt-outs apply to patients who have requested that their identifiable data is not shared outside of GP practices for purposes beyond their individual care and treatment.

The new service will follow a “collect once, use many times” method and harness the benefits of automation to do so. Initially, the collection will exclude records for patients who have exercised a Type 1 opt-out.

As this data set will be record level, data collected under these Directions may also be linked to other data sets held by NHS Digital subject to appropriate governance and authorisation, to provide more detailed information, enhance existing publications, develop new publications, and to respond to requests. More information about the data sets and collections that NHS Digital holds and that may be used for linkage can be found on the NHS Digital Data Collections and Data Sets webpage.

The new service will also provide new capability and capacity to meet the increasing demands for general practice data to support health and social care purposes including but not limited to planning, commissioning, policy development, public health (including COVID-19 purposes as described in the COVID-19 Public Health Directions 2020) and research on regional and national bases which cannot currently be met.

Overview of work being directed

NHS Digital is directed to replace the GPES extractions with General Practice Data for Planning and Research to provide the strategic collection and provision of general practice data for health and social care purposes including but not limited to planning, commissioning, policy, public health (including COVID-19) purposes and research. The Directions also support the implementation of the

recommendations made by the Public Accounts Committee as a result of the National Audit Office review of GPES.

GP system suppliers will perform the daily extraction and submission of General Practice Data for Planning and Research data collection from General Practices and other organisations as specified in the Scope below. GP system suppliers will pseudonymise the data at source using the national pseudonymisation tool provided by NHS Digital then securely transfer the data collection as a fully encrypted XML file to NHS Digital. NHS Digital will conduct checks to verify the completeness and accuracy of data.

It is intended to commence the collection before the end of March 2021.

NHS Digital will hold the collected data securely and may analyse and allow access to or disseminate the data in support of the purpose of the Directions in accordance with its legal powers, including its powers to allow access to or disseminate data under section 261 of the Health and Social Care Act 2012 (**the 2012 Act**) and any other legislation. NHS Digital will initially only publish data where either this has been agreed with the Department of Health and Social Care first, or where the Chief Statistician of NHS Digital considers this would be in accordance with the Code of Practice for Statistics, and in either case, this does not to any extent interfere with the performance by NHS Digital of its other functions in connection with this directed work or more generally.

The diagram below illustrates the high-level view of the data flow and changes in data controllership from before submission to receipt of disseminated data.

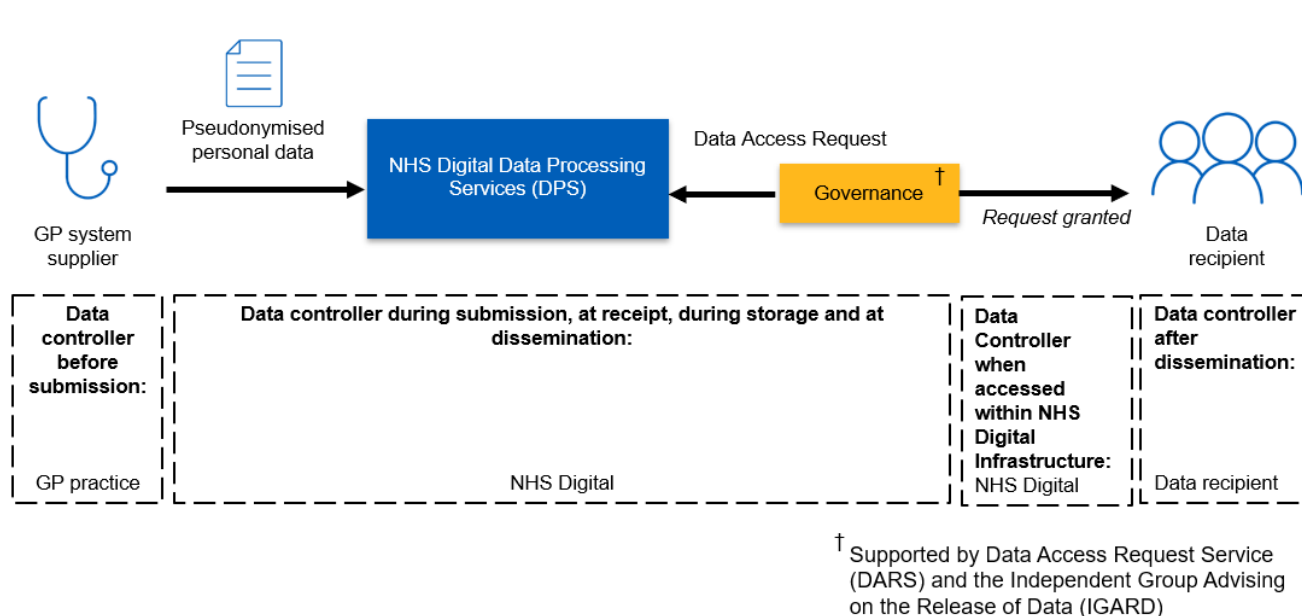


Figure 1 High level view of data flow and controllership

Data Collection

Scope

The Department of Health and Social Care (DHSC) requires NHS Digital to collect pseudonymised General Practice Data for Planning and Research data collection on a daily basis from the health and social care bodies listed below:

- General practices in England

Any changes to the patient record will be sent to NHS Digital to create a once daily data collection. A Data Provision Notice will be issued to all general practices and health and social care bodies as detailed above in England.

Source

GP system suppliers will extract the data identified in Annex B (General Practice Data for Planning and Research Data Set Specification) from the GP primary care systems of general practices and other instances of GP systems in England as set out in the Scope above, then prepare, pseudonymise and submit the pseudonymised GP Data Set (to NHS Digital. This pseudonymised data set is referred to as the GP Data Collection throughout this Annex.

Category of Data collected

Data is collected at patient level. NHS number, local patient ID, postcode, date of birth and date of death will be pseudonymised in the GP systems by the GP system suppliers before reaching NHS Digital. The GP Data Collection will be treated as personal data because NHS Digital will have the potential to re-identify this pseudonymised data upon the appropriate approvals as it will control the pseudonymisation keys. This is explained further below.

Special categories of personal data (as defined in the General Data Protection Regulation (GDPR), referred to as Special Category Data will be collected about patients in pseudonymised form relating to:

- Physical / Mental Health or Condition
- Sexual Life / Orientation
- Family / Lifestyle / Social Circumstance
- Religion or Other Beliefs
- Racial / Ethnic Origin

This data is coded and structured within, but not limited to, the following groups in GP clinical record systems:

- Patient Demographics
- Diagnoses and Symptoms
- Observations and Encounters
- Test Results
- Staff Details
- Medications, Allergies and Immunisations
- Referrals and Recalls
- Appointments

Legally restricted codes for Gender Recognition, Human Fertilisation and Embryology will not be collected.

Staff data is collected to identify the practitioner or other health professional involved in care. As these data items are already in the public domain, they are classed as personal data but not Special Category Data, therefore it is not necessary to pseudonymise this data.

Annex B General Practice Data for Planning and Research Data Set Specification provides details of each data item within the following worksheets: organisation-table, practitioner-table, patient-table, encounter-table, observation-table, immunisation-table, recall-table, allergy-table, medication-table, referral-table, problem-table, problem-link-table and appointment-table.

Analysis

Summary

General Practice Data for Planning and Research establishes an ongoing process for the collection, analysis by NHS Digital's data analytics function and controlled access to or dissemination of general practice data for health and care planning, commissioning, policy development, public health (including COVID-19) purposes and research.

Analysis by NHS Digital includes analysis and linkage of the GP Data Collection for data quality purposes, where NHS Digital can analyse data using a range of analytical tools to identify and address data quality issues across the collected data and other NHS Digital existing and new data sets. This could include undertaking comparison and consistency analysis of the same or similar information obtained in other collections, taking such steps as NHS Digital considers appropriate to maintain data quality, accuracy and consistency of information and address inconsistencies and inaccuracies across multiple datasets.

Where appropriate, analysis may also be carried out by NHS Digital, including linkage with other datasets, to produce aggregated data for dashboards and publication, to identify trends and generate insights, or to contribute to appropriate research and to create new linked data assets to support planning, commissioning, policy development, public health purposes (including COVID-19) and research.

This follows the 'collect once, use many times' approach, where data from single extraction can be used for multiple purposes. This means that less ongoing development work is required, and it reduces the burden on GP practices and GP system suppliers.

The pseudonymised data enhances patient privacy by referencing a person's records in a consistent way without identifying them. This will mean that when the right legal basis, controls and safeguards are in place, data can still be linked across different care settings and geographic boundaries, without revealing the identity of the patient.

The initial pseudonymisation process and process for re-identification are described in [System Delivery Function](#) below.

Type 1 Opt-Out and National Data Opt-Out

A Type 1 opt-out is an objection registered by a patient with their GP practice requesting the GP practice not to share the individual's personal identifiable confidential information outside of their GP practice except when it is being used for the purposes of their direct care.

Direct care is defined in the [Information Governance Review 2013](#) as "A clinical, social or public health activity concerned with the prevention, investigation and treatment of illness and the alleviation of suffering of individuals. It includes supporting individuals' ability to function and improve their participation in life and society. It includes the assurance of safe and high-quality care and treatment through local audit, the management of untoward or adverse incidents, person satisfaction including measurement of outcomes undertaken by one or more registered and regulated health or social care professionals and their team with whom the individual has a legitimate relationship for their care".

NHS Digital will not collect patient records which contain Type 1 opt-outs unless and until either:

1. Type 1 opt-outs are retired from use; or
2. NHS Digital agrees with the British Medical Association (BMA) and the Royal College of General Practitioners (RCGP), and the Department of Health and Social Care that it has put in place appropriate organisational and technical measures and controls in place to enable it collect and process pseudonymised Type 1 opt-out records by means which continue to uphold the Type 1 opt-out and do not enable the patient to be identified

In either case, this Requirements Specification will be updated to reflect these changes and in the case of point 2 above, to reflect the agreed organisational and technical measures and controls.

If data has already been collected by NHS Digital before a Type 1 opt-out was registered, the data collected before this will continue to be accessed and utilised as before. However, no further data will be collected from this record.

The national data opt-out will not apply to the collection of the data, as this is a collection which is required by law. It will be considered on access to or dissemination of the data in accordance with published national data opt-out policy.

Data Minimisation

Following discussions with the BMA and the RCGP, NHS Digital has identified data minimisation principles as listed below. This is to ensure that NHS Digital only extracts and holds the data that it needs in the GP Data Collection.

The fundamental principles and approaches adopted in the data minimisation are:

1. GP system suppliers will pseudonymise NHS number, local patient ID, postcode, date of birth and date of death before providing the data to NHS Digital
2. GP system suppliers will create and append postcode sector, Lower Layer Super Output Area (LSOA)¹, week/year of death and week/year of birth before providing the data to NHS Digital. This reduces unnecessary processing by making these separately available for granular analysis where legal and approved
3. Only structured and clinically coded data will be collected (free text, images and documents will not be collected)
4. Legally restricted codes for Gender Recognition and Human Fertilisation and Embryology will not be collected
5. Certain data items in certain categories will be subject to limitations on the age of the data and will not be collected. Such limitations will apply where historic data is not required to meet existing uses due to its age (e.g. a 10-year limit on medications)
6. Data will only be collected for patients with registration types set as 'GMS Registered', 'GMS Applied' and 'Temporary Registration'

Data collected in the GP Data Collection will be subject to ongoing audit of actual data use to compare what data is approved for specific uses against what data NHS Digital is collecting. Where there is clear evidence that data is being collected which is not required, this data will be removed from the collection and subsequently deleted from NHS Digital secure storage in accordance with NHS Digital's Records Management Policy.

Data Flow

The GP Data Collection will provide a comprehensive and frequent (daily) extract of patient level and linkable data from the GP system suppliers. All identifiable data items will be pseudonymised by the GP system suppliers and flow into NHS Digital's Data Processing Services platform (DPS).

Once the pseudonymised GP Data Collection is received in DPS it is stored securely. The data is then made available for appropriate and approved internal NHS Digital access and to be linked to other data collections, subject to the relevant Internal Analytics Access Approval Process or approval processes for controlled access to or dissemination of the data via the Data Access Request Service (DARS) as set out below.

The table below describes the data processing and data flow. All data is transmitted and stored in encrypted form.

¹ LSOA represents a small region or cluster of post codes and does not identify the specific post codes

Table 1: GP Data for Planning and Research Data Processing and Flow

Category and Description of Processing	Entity involved	Types of Personal Data Processed	Data Processing and Flow
Stage 1: Authorise the extraction of data from the practice clinical system	Data Controller: GP practice Data Processor: GP system supplier	Identifiable patient data	Data is held in GP patient record systems which meet national standards for security under relevant GP IT system contracts.
Stage 2: Pseudonymisation process in practice clinical system.	Data Controller: GP practice Data Processor: GP system supplier	Pseudonymised patient data	Identifiable patient data will be pseudonymised by GP system supplier on behalf of the GP practice using the national pseudonymisation tool provided by NHS Digital.
Stage 3: GP system suppliers provide data to NHS Digital	Data Controller: GP Practice Data Processor: GP system supplier	Pseudonymised patient data	Pseudonymised data will be transferred by GP system supplier to NHS Digital through an approved secure method (either Message Exchange for Social Care and Health (MESH) or Amazon Web Services (AWS)). In the event that the transfer of data to NHS Digital fails, then the transfer will be repeated. In the event that the GP system supplier is unable to pseudonymise the data, there will be no data transferred to NHS Digital by the GP practices contracted to the relevant GP supplier.
Stage 4: NHS Digital receives data into the Data Processing Service platform	Data Controller: NHS Digital	Pseudonymised patient data	Data is held in pseudonymised form in the Data Processing Service platform (DPS) of NHS Digital

Category and Description of Processing	Entity involved	Types of Personal Data Processed	Data Processing and Flow
Stage 5: NHS Digital checks data for completeness	Data Controller: NHS Digital	Pseudonymised patient data	Data is processed in DPS to verify completeness and accuracy of data In the event that completeness and validation checks fail, then the cause will be investigated and, when resolved, the process will be restarted from stage 1 with the regeneration of the data extract. Processing of the data will not proceed to Stage 5 until the data quality issues are resolved.
Stage 6: Data is made available to end users	Data Controller: NHS Digital Data Controller: Data Recipients	Pseudonymised patient data or, subject to appropriate approvals and re-identification procedure, identifiable patient data	NHS Digital data analysts can access and analyse the pseudonymised data in accordance with the NHS Digital Internal Analytics Access Approval Process described below. Data recipients can access the pseudonymised data following approval via the DARS process for secondary use. Where access by internal NHS Digital Analysts or data recipients is required to identifiable data, this will be subject to prior approval in accordance with the NHS Digital Re-Identification Approval Process. This Approval Process requires approval from the Privacy, Transparency and Ethics Team, the Chief Medical Officer (CMO), the Senior Information Risk Officer (SIRO) and the Caldicott

NHS Digital Internal Analytics Access Approval Process

When access to data and/or linkage is required for a specific analysis internal to NHS Digital then the NHS Digital Internal Analytics Access Approval Process will apply to control internal access for analysis which will include approval by the Privacy, Transparency and Ethics Team and Caldicott Guardian Team. Analysts will only be granted access to those functions or data required to carry out approved internal analysis.

Access to identifiable data for internal analysis will only be granted where there is a specific need that cannot be met using the pseudonymised data. This will be assessed as part of the Internal Analytics Access Approval process and will be subject to additional approval through the Re-Identification Approval Process. This process ensures that authorisation is only given to analysts for a time-limited period, where the access to the data is justified, and an appropriate legal basis for such processing is in place.

Consultation

NHS Digital, in collaboration with the Department of Health and Social Care (on behalf of the Secretary of State) aims to build informed and clear public and professional support for the use of general practice patient data for planning and research and has consulted with a range of organisations and stakeholder groups to help achieve this aim as required by section 258 of the Health and Social Care Act 2012. In particular NHS Digital has engaged with representatives of general (medical) practitioners (GPs) through the BMA and RCGP for the three years leading up to the issue of the Directions.

Principal engagement has been through the Joint GP IT Committee (JGPITC - a sub-group of the RCGP formed of GPs with specialist interests in data and IT and policy leads and ethics representatives from the BMA). The Clinical Lead and other senior members of NHS Digital's General Practice Data for Planning and Research team attend quarterly JGPITC meetings to provide an update and invite feedback in addition to separate dedicated sessions of co-design and challenge.

NHS Digital also commissioned an independent review of the data collection proposal and oversight of methodology by the Faculty of Clinical Informatics (FCI) at the request of the JGPITC in 2019 to provide independent assurance. From this a series of recommendations were made which NHS Digital considered and responded to, both with the FCI and the profession. NHS Digital continues to work with the professional representatives to build ongoing support and confidence.

NHS Digital has engaged with stakeholders on an ongoing basis and their feedback and advice have also informed the development of other critical work areas for General Practice Data for Planning and Research, including:

- The Privacy, Transparency and Ethics Team, the Caldicott Guardian and the Data Protection Officer – Ensuring data is protected and processed in line with national policy, data protection law and the Caldicott Principles
- Business processes – Developed and implemented in line with the needs of stakeholders.
- Technical architecture – Ensuring the collection and processing of data is aligned to information governance, business processes and the Data Processing Services (DPS) programme's requirements
- Consultation has taken place with a number of stakeholders including those who represent patient and public interest in data - for example, the Office of the National Data Guardian and Healthwatch England

Organisations and stakeholder groups consulted with include:

- The BMA
- The RCGP
- The Department of Health and Social Care
- NHSX
- NHS England and NHS Improvement
- The National Data Guardian
- Clinical Commissioning Groups, Commissioning Support Units and Data Services for Commissioners Regional Offices
- Research bodies including Public Health England, Health Data Research UK, Healthcare Quality Improvement Partnership, Genomics England, various university institutions

- The Data Coordination Board (DCB)², and the newly formed Data Alliance Partnership Board (DABP) which has now replaced DCB, which includes representatives from the Department of Health and Social Care, The National Institute for Health and Care Excellence, NHS England and NHS Improvement, Public Health England, Care Quality Commission, Local Government Association, Health Education England, Health Research Authority, Association of Directors of Adult Social Services and NHS Digital
- The Data Alliance Partnership Board
- The Information Commissioner's Office
- The Independent Group Advising on Release of Data (IGARD)
- Representatives of patients including Understanding Patient Data, Healthwatch England, Health Data Research UK patient and public involvement and engagement group, Use My Data and a small sample of GP Patient Participation Groups
- GP system suppliers contracted to supply systems via the GP IT Futures Lot 1 Framework Agreement

Data Access and Dissemination

GPES disseminations

The GP Data collection will replace the existing regular GPES disseminations with data collected on a phased transition basis, commencing with those disseminations which are not dependent on inclusion of data from GP medical records with Type 1 opt-outs.

–Data access via the NHS Digital Data Access Request Service (DARS)

Access to GP data for secondary use purposes by third parties will be subject to approval through the Data Access Request Service (DARS) with appropriate oversight from the Independent Group Advising on the Release of Data (IGARD) and the Professional Advisory Group, where new, novel or identifiable data applications are being considered.

- If approved, the data will either be made available for access within a secure NHS Digital data access environment or, where the needs of the recipient cannot be met in this way, through a direct dissemination
- All data access or release for planning or research purposes will be covered by binding Data Access Terms or a Data Sharing Agreement

Publication

Data prohibited from being published

NHS Digital must not publish any information obtained as a result of complying with the Direction unless this is first agreed with the Department for Health and Social Care or where the Chief Statistician of NHS Digital considers this would be in accordance with the Code of Practice for Statistics and in either case, this does not to any extent interfere with the performance by NHS Digital of its other functions in connection with GP Data for Planning and Research or its functions more generally.

² The Data Coordination Board (DCB) forms part of a system-wide information and technology governance model. This has now been replaced by the Data Alliance Partnership Board (DAPB). The DCB (now DAPB) acts with delegated authority from the Secretary of State as the main governance route through which all information standards, data collections and data extraction requirements for use in health and care in England will be agreed and priorities assigned.

System Delivery Function

In accordance with Regulation 32 of the Regulations, the Secretary of State directs NHS Digital to exercise such systems delivery functions of the Secretary of State as are necessary for it to deliver the **General Practice Data for Planning and Research Directions Information System**.

The General Practice Data for Planning and Research data collection will utilise NHS Digital's Data Processing Services (DPS). This is web-based portal of secure technology for data collection, processing and data access that is transforming the way data is used to improve health and care. It enables NHS Digital to collect, process and make data available for use in a smarter, more secure and efficient way and deliver more timely and better linked data to give a clearer national picture of health and care. These secure tools and processes will support future innovation and improvements.

Pseudonymisation and linkage

The General Practice Data for Planning and Research data collection service uses the national pseudonymisation tool procured by NHS Digital. This is a third-party tool that is supported by robust security accreditations and assurance processes. Its underlying technology includes a form of encryption which allows for the generation of consistent pseudonyms for the safe transfer and linkage of data without the need to reidentify. This use of homomorphic encryption means that the original identifiers once transformed can be processed and further pseudonymised without loss of information or need to re-identify.

The national pseudonymisation tool is provided to each of the GP system suppliers who install this within their own infrastructure. Each GP system supplier will then prepare the extract and once the extract file is ready for transfer, use the pseudonymisation tool to pseudonymise the relevant patient identifiers (NHS number, local patient ID, postcode, date of birth and date of death). The tool then creates the unique pseudonym (transit pseudonym or pseudonymised ID) for each patient. The GP system supplier then further processes the data extract and securely transfers the data collection as a fully encrypted XML file to NHS Digital via either the Message Exchange for Social Care and Health (MESH) or Amazon Web Services (AWS).

When the extract files are received by NHS Digital, they are processed and stored in their pseudonymised state. These extract files are checked for completeness as part of the data quality processing and are then stored independently from all other data collections.

Whenever data from the GP Data Collection is requested for a use or requires linkage to another dataset then, following approval via the established governance processes, the relevant authorised data management team will create a new view of the approved data items within the DPS platform. This view will take the relevant identifiable data items from the other dataset and pass them through the national tool to provide a linkable pseudonymisation token. The data management team will then take the pseudonymisation token from the GP Data Collection, transform this again, and using the pseudonymisation tool, perform a link that ensures that the pseudonymised NHS numbers from the two data sets are correctly matched. This linkage is performed in a new domain within the NHS Digital boundary, separate to both of the original data sets. Each time the tool is used to perform a link between two data sets it creates a new pseudonym ensuring that the original pseudonym is protected. This means that if pseudonymised data is required in five different, approved uses, then the pseudonymised identifier will be different in all five collection outputs.

Re-Identification and controls

When data is approved for release in identifiable form following the Re-identification Approval Process, the national pseudonymisation tool is able to reverse the pseudonymisation process by use of a re-identification key that is inbuilt within the national tool. Only a small number of high security administrators within NHS Digital have credentials to perform this process and have the relevant role-based access controls which enable them to execute the re-identification procedure. They will only execute the re-identification procedure following approval through the Re-identification Approval Process.

All approved, re-identification requests, will transform the relevant pseudonyms to a clear value (e.g. NHS number) and will be made available to the data recipient. This will apply to external and internal requests.

For data to be re-identified, the de-identification service will take the relevant pseudonymised fields and map these back to the Root ID (this is the original pseudonymisation token applied once the data lands and is transformed from the transit ID applied by the GP system supplier). Once the reversal of the pseudonym to the Root ID is complete the tool then performs a secondary process transforming the Root ID into the encrypted identifier (transit ID) which is then decrypted to the original identifier, this clear identifier is then securely made available to the data recipient.

The Root ID is never made available to data recipients. The encrypted identifier known as a Transit ID is a private key known only to NHS Digital.

The original pseudonym is not shared with data recipients nor are the Root IDs. Only the randomised pseudonym or, where approved in accordance with the Re-Identification Approval Process, identifiable versions of pseudonymised items are ever shared.

Change control process

Changes to this Specification will be managed in conjunction with the Department of Health and Social Care.