York Teaching Hospital
NHS Foundation Trust

Anonymisation of Data (Pseudonymisation) Policy and Procedure

Author:	Fiona Jamieson
Owner:	IGEG
Publisher:	Healthcare Governance Unit
Version:	1
Date of version issue:	12/3/2019
Approved by:	IGEG
Date approved:	12/3/2019
Review date:	April 2021
Target audience:	
Relevant Regulations and Standards	DSP Toolkit Requirements
Links to Organisational/Service Objectives, business plans or strategies	

Executive Summary

The purpose of this policy and procedure is to advise staff of the purpose and requirements of using pseudonymising data instead of patient identifiable data for purposes that are not directly healthcare medical related.

The implementation of these procedures will mitigate the risks of breaching the Data Protection Act 1998 by allowing staff to view and use patient identifiable data when there is no need for them to do so.

Version History Log

This area should detail the version history for this document. It should detail the key elements of the changes to the versions.

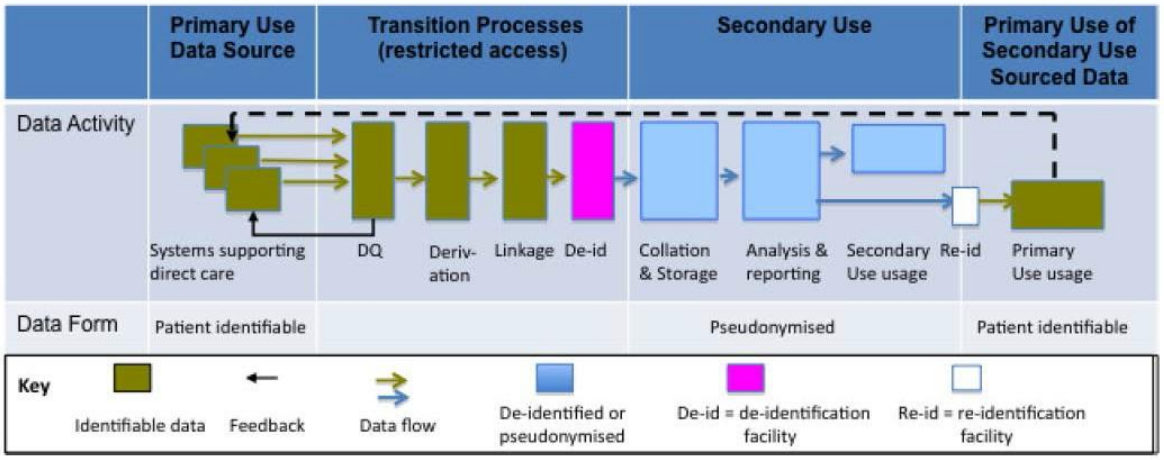
Version	Date Approved	Version Author	Status & location	Details of significant changes
1	12/3/2019	Deputy Director of Healthcare Governance		First policy of its type within the organisation aiming to formalize the processes that have been in place for some time.

Contents

1 Introduction & Purpose

The purpose of this policy and procedure is to advise staff of the purpose and requirements of using pseudonymising data instead of patient identifiable data for purposes that are not directly healthcare medical related.

The implementation of these procedures will mitigate the risks of breaching the Data Protection Act 1998 by allowing staff to view and use patient identifiable data when there is no need for them to do so.



The overall aim of pseudonymisation is to enable the legal, safe and secure use of patient data for secondary (non-direct care) purposes by the NHS (and other organisations involved in the commissioning and provision of NHS- commissioned care) and to enable NHS businesses to no longer use identifiable data in its non-direct care related work wherever possible.

2 Scope & Definitions Scope

This document applies to all directly and indirectly employed staff within York Teaching Hospitals NHS Foundation Trust and other persons working within the organisation in line with York Teaching Hospitals NHS Foundation Trust Trust's Equal Document. This document is also recommended to Independent Contractors as good practice

2.1 DEFINITIONS

Personal Identifiable Data (PID)

Any single data item, e.g. NHS number or group of data, e.g. first name, last name, postcode and date of birth, which can be used to identify an individual patient. In some cases where a small group is involved, e.g. HIV patients, a postcode might be sufficient to identify a patient and so can be considered patient identifiable.

Primary Uses (Healthcare Medical Purposes)

The primary use of data is any purpose that directly contributes to the safe care of the patient. This includes:

- Care
- Diagnosis
- Referral and treatment processes
- Drug safety

- Public health surveillance
- Relevant supporting administrative processes such as:
    - Clinical letters
    - Patient administration
    - Patient management on a ward
    - Managing appointments for care
    - Audit/assurance of the quality of healthcare provided

Secondary Uses (Non-Healthcare Medical Purposes)

A secondary use of data is any use which is not covered in the definition of a primary use. In essence it relates to the use of patient identifiable information which does not directly contribute to the safe care of the individual concerned. Examples of secondary use of patient data include performance management, commissioning and contract monitoring.

De-identifying Data

The process of using one or more techniques that transform data to make it less likely that individuals can be identified. The goal of de-identifying data is to render it "effectively anonymised." De-identification techniques include stripping out person identifiers, pseudonymisation, aggregation and derivation. Section 9 details how these processes should be used in practise.

Effectively Anonymised Data

Data is "effectively anonymised" when the recipient is unable to infer the identity of individuals from the data without the application of effort or resource where it would be unreasonable to anticipate in the circumstances that apply.

Effectively anonymised data would almost certainly neither be considered "personal data" nor "sensitive personal data" under the Data Protection Act 1998, nor "confidential patient information" under the NHS Code of Confidentiality 2006.

Stripping Out Person Identifiers

This is the process of removing person identifiers from data. This may be partial (where only some identifiers are removed) or complete.

Pseudonymisation

The process of replacing person identifiers in a dataset with other values (pseudonyms) from which the identities of individuals cannot be intrinsically inferred. Examples of this process are replacing an NHS number with another random number, replacing a name with a code or replacing an address with a location code.

Pseudonyms themselves should not contain any information that could identify the individual to which they relate (e.g. should not be made up of characters from the date of birth, etc.). The correct application of this process will produce the same pseudonym for a patient across different data sets and time so that patient data can still be linked.

Aggregation

This is the process of pooling data such that category totals are displayed rather than individual values. Care must be taken so that when small datasets are used an individual's identity cannot be inferred because they are the only person in a

category. Most reports for contract and performance purposes provide aggregated data.

Derivation

This is the process of creating one piece of new information from original data. For example, the ward or residential Clinical Commissioning Group (CCG) of a patient is derived from the patient's full postcode.

The aim of using derivations is to display values that reflect the character of the source data, but which hide the exact original values. This is usually done by using coarser-grained descriptions of values than in the source dataset e.g. replacing dates of birth by ages or years, addresses by areas of residence or wards, using partial postcodes, rounding exact figures so they appear in a normalised form.

When original values are replaced by a range (for example, date of birth replaced by an age range) this is known as banding.

Data Quarantining

The technique of only supplying data to a recipient who is unlikely to have knowledge of the data e.g. part of a dataset to a recipient, so they do not know from which part of the country the data has emanated (i.e. it contains no local data) or providing data to a recipient who does not know the clinical domain to which the data relates, or providing data with a local patient identifier attached, the meaning of which is not available to the recipient.

New "Safe Haven"

A new "Safe Haven" refers to a back office function whereby authorised individuals are allowed access to patient identifiable information so that data quality checks, derivation, pseudonymisation and accurate record linkage for the same patient can be undertaken. Restriction to the data is achieved by access control measures meaning that a new safe haven is virtual in nature rather than physical.

Within York Teaching Hospitals NHS Foundation Trust[1], the new "Safe Haven" for patient identifiable information for secondary use purposes is the Information Team.

Access Control

A system whereby access to data is restricted to appropriate staff based on individual login names and/or membership of security groups. Examples of

this include Role Based Access Control (RBAC) for smartcards and network drive security.

Business Process Types

There are three main types of business process or component steps within an overall business processes that need to be considered:

- A - those processes using patient data involved in the direct care of patients, e.g. patient administration, such as booking appointments, managing waiting lists etc.;
- B - those processes using patient data not involved in the direct care of patients, e.g. analysis of waiting lists or monitoring referral-to-treatment times;

[1]

- C – a combination of Types A and B; here the purpose of the overall business process is to process and analyse patient records in order to understand and report on specific issues, which may subsequently require interaction with patients, e.g. the use of the PARR algorithm.

3 Process/Requirements

3.1 Secondary Purpose Exemptions

In the following circumstances, identifiable data can be used for secondary purposes:

- The patient's consent
- Legal requirements, such as the Mental Health census
- Regulations relating to specific organisations and their function, such as the Care Quality Commission, Audit Commission and Health Protection Agency.
- Regulations under Section 251 relating to health functions, such as Cancer Registries, communicable diseases and other Public Health functions.
- Approval by the Secretary of State for specific or class approval under Section 251, such as research projects.

3.2 Spatial Analysis

The spatial analysis of patient level data is a type B purpose (not for direct patient care) which is important for identifying and addressing concentrations of health problems and determining the access that patients have to healthcare facilities and services.

Robust spatial analysis can only be undertaken if the full postcode of a patient or service user is available. As postcode is an important potential contributor to identification of patients, it is necessary to make data and analytical facilities available to end-users without revealing postcodes of individual patients.

Analysis should therefore be undertaken either:

- By the end-user using data that has been effectively anonymised including derived areas but no postcodes, or
- Within the New Safe Haven by an Information Analyst. Output should be provided in suitable mapped plots and if patient level data is required by end users, then it should be provided in pseudonymised form with modified postcodes, such as postcode sector or blurred forms.

3.3 Dealing with Type C Processes

Sometimes the analysis of data can be for both primary and secondary care purposes, e.g. identifying patients at risk of readmission .The initial analysis is for secondary purposes but the resultant output can be used for direct patient care.

The analysis can be undertaken in one of two ways:

- A clinician can initiate/undertake the analysis themselves using standard analyses and reports on the basis of legitimate relationships with the patients;
- The analysis can be undertaken as a secondary use with de-identified data and for the selected cohort to be made available to the relevant clinicians in identified form.

3.4 Inter-Organisational Communications

The following policies and procedures only relate to patient data used for secondary purposes and not when patient care is involved, i.e. sharing information between York Teaching Hospitals NHS Foundation Trust and the patient's GP or consultant.

Identifiable Patient Data - This type of data may only flow from either the service or the New Safe Haven (Information Services) to the other organisation's New Safe Haven. For example, invoices raised by finance for out of area treatment which are sent to the commissioning officer must not contain patient identifiable information.

Commissioned Data Set (CDS) and Minimum Data Set (MDS) Extracts - Section 251 allows for the submission of patient identifiable information to be included in patient activity data submitted to SUS (Secondary Uses Service) as a CDS extract. No section 251 agreement exists for the direct submission of PID data to commissioners.

Paper-based Communications - With immediate effect no paper-based communication of data for secondary use purposes must be made, even if done via New Safe Havens. The reason for this is to ensure that the data is transmitted more securely and in an encrypted form to reduce the chance of some not authorised to see the information viewing it or it being lost or intercepted in transit.

Queries Relating to Activity - For communications between organisations relating to activity the patient's NHS number must not be used. Instead the service's episode, referral ID or local patient identifier must be used.

Wherever possible an episode or referral identifier should be used in preference to a local patient identifier created by the clinical system used by the service in question.

Queries Relating to Data Quality - Where communication is required in connection with data quality then local identifiers or NHS numbers can be used as long as they are transmitted between the commissioner's New Safe Haven and either the Trust's New Safe Haven or to the service in question.

Invoicing - Verification of patients for invoicing must not contain patient identifiable information such as NHS numbers, dates of birth, postcodes etc. To enable commissioner to verify that the patients invoiced for are their responsibility, GP practice code information should be supplied.

In circumstances where care is delivered jointly by York Teaching Hospitals NHS Foundation Trust and commissioner staff, for example Intermediate Care or Substance Misuse, there may be requests for patient identifiers so funding can be tracked for patients between providers. In such cases, and where staff from both organisations use the same clinical system, such a request can be accommodated by supplying a local system patient ID as part of the invoice backing data. Such a reference can only be used by staff with a direct relationship with the patient to identify them and thus mitigates finance and other admin staff from having access to patient identifiable data.

Patient Surveys – These are classified as service evaluation or research and not as clinical audit. As such, patient surveys are a secondary use of patient data if the cohort of people being surveyed is chosen from patient based records, hence care needs to be taken in order to avoid breaching patients' confidentiality.

Confidentiality may be breached not only through disclosure of identifiable data to a third party employed to undertake the patient survey but also by inference; for

example, if the people to be sent the survey have been selected on the basis of clinical criteria, from which the third party could deduce that the listed people had a particular condition or had received a particular procedure or simply that they had had contact with services. The central question in determining whether confidentiality will be breached is which organisation undertakes the selection and administration of the survey. If the cohort is selected by an external agency or if the hospital selects but then provides the details to a data processor to send out the letters then because there are clinical criteria e.g. inpatient in the last six months, then this would be classed as a disclosure and, if there were not prior consent from the patient, would be a breach of confidence.

Examples of good information governance practice are:

- If York Teaching Hospitals NHS Foundation Trust selects its own patients and sends out the survey with the responses going to a company for analysis then, provided this is clear to the patients, there is consent from the patient returning the survey and no breach of confidence would occur.
- If there is prior consent to allow disclosure for patient surveys to be sent, then again no breach of confidence would occur.
- If the above are genuinely not practicable there are two alternative approaches:
    - Obscure the clinical inference e.g. people receiving treatment in the last six months but adding a substantial proportion (at least 20%) that have been randomly selected based only on demographic information

      e.g. living in the same area and in the same age categories.
    - Create two lists, one with a patient ID (pseudonym), name and address, i.e. the mailing list, and the other with the same patient ID and the service used, i.e. the analysis list. The company used to conduct the survey must be contractually required to keep the lists separate.

      The people responsible for sending out the questionnaire may only have sight of the mailing list. The people responsible for the analysis should only have access to the analysis list. This process minimises the risk of a third party being able to infer from the data supplied which service a patient has used.

Anyone considering conducting a patient survey should seek guidance from Information Governance and Patient and Public Experience Co-ordinator on this matter first to ensure that confidentiality issues are fully considered and addressed prior to patient contact.

3.5 De-identification Methods

Sensitive Data Definition - The following data items are considered sensitive and should be subject to de-identification processes when data is to be used internally for secondary use purposes. They can relate to a patient, baby or mother depending on the nature of the information.

- Name
- Address
- Date of birth
- Postcode
- NHS Number

- Ethnic category
- Local patient identifier
- Hospital Spell Number
- Patient Pathway identifier
- Secondary Uses Service (SUS) Spell Identifier
- Unique Booking Reference Number
- Social Services Client Identifier
- Date of death

Appendix 1 shows the recommended way of handling each data field type to effectively anonymise the data.

Means of De-identification - De-identification of patient records can be achieved through all or a combination of:

- Not displaying sensitive data items
- Using derivations to replace the values of certain data items in systematic ways, such as using:
    - electoral ward instead of postcode, displaying age instead of date of birth
    - banding of values, such as displaying age bands (e.g. 5-10) instead of date or year of birth
    - using post code sector (first 4 characters e.g. DE3 7) instead of the full post code e.g. DE3 7FZ
- Using pseudonyms on a one-off basis
- Using pseudonyms on a consistent basis

Surrogates - A surrogate is a substitute for another entity, such as a data item. A pseudonym can be considered as a substitute data item (or surrogate).

If surrogates are used as a basis for de-identification, these surrogates must be:

- Unique system-wide (e.g. replacing NHS Number), hence never reused
- System generated (i.e. not created by the user)
- Not manipulable by the user or application
- Without semantic or obvious meaning
- Not composed of several values from different domains.

The use of a surrogate field as a pseudonym for an identifier such as the NHS Number, may require a table being maintained giving a one-to-one correspondence between the surrogate and the identifier.

3.6 Creation of Pseudonyms

Appendix 2 provides a diagram of the pseudonymisation process and also includes rules and best practice recommendations for its implementation.

3.7 Provider Stand-Alone and Legacy Systems

There are two sets of circumstances where de-identification of data for secondary uses within York Teaching Hospitals NHS Foundation Trust may not be immediately feasible. These are where the addition of pseudonymisation facilities or the modification of reporting functionality to existing systems using patient identifiable

data would be complex and disproportionately uneconomic.

Such systems are:

- stand-alone systems, usually supporting specific clinical areas
- legacy systems, which may include Patient Administration Systems (PAS).

For many of the systems which fall into this category there is already an implemented solution. Data is exported from the clinical system into the New Safe Haven data warehouse and then reports are generated which supply the information required by the service via a web-based report manager. As reports are generated by database queries no patient identifiable data is revealed unless it is needed for primary use purposes. Access to reports is controlled using Windows login permissions.

Where such an approach is not feasible due to time constraints or complexity then York Teaching Hospitals NHS Foundation Trust must mitigate the risk posed by these systems by:

- Compiling a register of such systems
- Implementing appropriate restrictions on access to identifiable data made available through these systems through physical and electronic means
- Training of relevant staff on secondary use information governance and good practice
- Developing exit strategies to resolve the information governance issues for each system

3.8 Access Control to Patient Identifiable Information

Clinical Systems - All clinical systems must have some form of access control, either by password or preferably by smartcard.

Information Asset Owners should ensure that access is only granted to those individuals that need to work with patient information and that this is restricted to the minimum level of access required to do their job.

New Safe Haven - All data held within or used by the New Safe Haven is stored in either the data warehouse or on the Information Services' shared network drive.

Access to the data warehouse is controlled by the Software Development team (part of Information Services) who are able to limit what data each team member is able to view in patient identifiable form.

Data held on the shared network drive can only be accessed by members of the Information Services team. The Head of Information Services has direct control over who has access to this area.

Authorising Access to Patient Identifiable Information for Secondary Use Purposes - As part of the Information Asset Register, an Access Control List (ACL) must be held of all individuals who have been granted permission to view patient identifiable information for secondary use purposes. This list must detail the name of the individual, what access they have been granted, justification as to why access is needed, the date authorisation was granted and the date authorisation was revoked.

Authorisation can only be granted by the Information Asset Owner, Caldicott Guardian or Senior Information Risk Officer.

In the case of the New Safe Haven authorisation can only be granted by the Caldicott

Guardian or Senior Information Risk Officer.

Access Review - On a quarterly basis the Access Control List should be reviewed to ensure it is up to date and that if any member of staff has left or changed job roles they are removed from the list and their access to data stopped if this is appropriate.

Data Access Logging - It is necessary to log access to identifiable data. This is in order to provide the basic information to support the Care Record Guarantee to inform patients as to who has accessed/seen their data and to support forensic analysis in the event of untoward incidents.

The logging process should happen automatically in transaction processing systems and would be expected for provider based clinical systems. However, for database systems primarily operated for non-direct care purposes, such as the data warehouse, this is not feasible as the records for legitimate accesses could be as numerous as the database itself and would also contain identifiable data, exacerbating the risk of inappropriate disclosure of patient identifiable data.

The aim of logging becomes to create the ability to know who has accessed identifiable data and to be able to replicate the query undertaken. The key items to be logged therefore are:

- Who has accessed which databases containing identifiable data
- Date and time of access
- Query or access process undertaken, including the parameters of the query.

The log of accesses should itself form a structured database to enable queries and audit.

Auditing - Periodically and no less than once per quarter, the Information Asset Owner or nominated person should review the log of accesses. This should be done as an audit via the sampling of users or subject matter. The aim of the audit is to check for unusual patterns of access.

A log should be held showing when the audit was conducted, any issues that were spotted and what was done about these issues. Any issues identified must also be reported using the Incident Reporting system.

Staff With Dual Roles - In some cases members of staff will have a legitimate need to access patient identifiable information as well as pseudonymised data as part of their job. For example a receptionist might book patient appointments and also run off reports for performance monitoring.

Where possible business processes should be changed so that there is a separation of duties. That is one member of staff has access to patient identifiable information but only has duties which require access to this data for primary uses, while another only has access to pseudonymised data.

Where such separation is not practical, authorisation for the individual to access both sets of information must be approved by the Caldicott Guardian or Senior Information Risk Officer.

4 Roles & Responsibilities

4.1 All Staff

It is the duty of all staff to:

- Understand the difference between primary uses (for healthcare medical purposes) and secondary uses (for all non-healthcare medical purposes) of information.
- To only use effectively anonymised data when working with patient information for secondary use purposes.
- To seek guidance from their line manager, the Head of Information Services or Information Governance if they are unsure about whether they should be using pseudonymised data.

4.2 Managers

All managers should ensure that:

- Their staff read, understand and implement where necessary the policy and procedures in this document.
- They review their business processes and where necessary and practical make changes to separate roles and responsibilities so that individuals aren't required to use both patient identifiable and anonymised information as part of their job.
- Submit Secondary Use Access Authorisation Request forms to the appropriate Information Asset Owner to add or remove individuals from the list of people authorised to use patient identifiable information for secondary use purposes. The number of people authorised should be kept to a minimum.

4.3 Information Asset Owners

Information Asset Owners are required to:

- Authorise requests for access to Information Assets they are responsible for and keep a register of such approval or non-approval.
- Ensure logs and audit trails are created and retained by the information systems of who has accessed patient identifiable information.
- Ensure that the logs are audited periodically and actions taken if issues are found.
- Ensure that IT systems have the ability to provide separate views of pseudonymised and identifiable data so that access can be granted to individuals to view the appropriate information they need to do their job.

4.4 Caldicott Guardian and Senior Information Risk Officer (SIRO)

The Caldicott Guardian and SIRO are required to:

- Authorise requests for access to Information Assets on behalf of the Information Asset Owner.
- Authorise requests for access to data held in the New Safe Haven for the purposes of de-identifying data

4.5 Head of Information Services

The Head of Information Services is responsible for:

- Keeping New Safe Haven identifiable and pseudonymised data logically separate.
- Ensuring that pseudonymisation functionality is implemented within the Solent data warehouse.

- Ensure that access to data is limited to those individuals who require it for de-identification purposes and matches that authorised by the Caldicott Guardian or SIRO.
- Finding and implementing solutions to mitigate the risks posed by stand- alone and legacy systems where patient identifiable and pseudonymised data cannot be provided and/or separated.

4.6 Information Governance

The Information Governance should:

- Assist Information Asset Owners with setting up Information Asset Registers
- Ensure Information Asset Owners understand their responsibilities
- Keep a register of stand-alone and legacy IT systems which do not provide and/or separate patient identifiable from pseudonymised data.

5 Training

Solent NHS Trust recognises the importance of appropriate training for staff. For training requirements and refresher frequencies in relation to this policy subject matter, please refer to the Training Needs Analysis (TNA) on the intranet.

6 Equality Impact Assessment and Mental Capacity

The outcome of the Impact Assessment (see appendix 4) was negative, i.e. this policy and process has no adverse impact on equality or mental capacity of those people affected by it.

7 Success Criteria / Monitoring Effectiveness

The table below outlines the Trusts' monitoring arrangements for this policy/document. The Trust reserves the right to commission additional work or change the monitoring arrangements to meet organisational needs.

Aspect of compliance or effectiveness being monitored	Monitoring method	Individual responsible for the monitoring	Frequency	Committee to receive monitoring report	Committee responsible for ensuring actions are completed
Compliance with pseudonymisation procedures and safe haven guidance	Audit of data production processes	Head of Information Services	Annual	IG Steering Committee	IG Steering Committee

Following the review outlined above, any non-compliance with this policy will be reported as part of the audit and if appropriate logged as an incident.

8 Review

This document may be reviewed at any time at the request of either at staff side or management, but will automatically be reviewed twelve months from initial approval and thereafter on a three tearly basis unless organisational changes, legislation, guidance or non-compliance prompt an earlier review.

9 References and Links to other Documents

The following documents are referred to in this policy:

- Data Protection Act 1998
- Human Rights Act 1998
- Health Service (Control of Patient Information) Regulations 2002
- Confidentiality: the NHS Code of Practice

The following documents were used as reference during the writing of this policy and procedure:

- Pseudonymisation Implementation Project (PIP): Summary of Pseudonymisation Implementation Guidance
- Pseudonymisation Implementation Project (PIP): Implementation Guidance on Local NHS Data Usage and Governance for Secondary Uses
- Pseudonymisation Implementation Project (PIP): Reference Paper 1 – Guidance on Terminology
- Pseudonymisation Implementation Project (PIP): Reference Paper 2 – Guidance on Business Processes and Safe Havens
- Pseudonymisation Implementation Project (PIP): Reference Paper 3 – Guidance on De-identification
- Pseudonymisation Implementation Project (PIP): Reference Paper 4 – Pseudonymisation Technical White Paper – Design and MS-SQL
- Incident Reporting Policy

10 Glossary

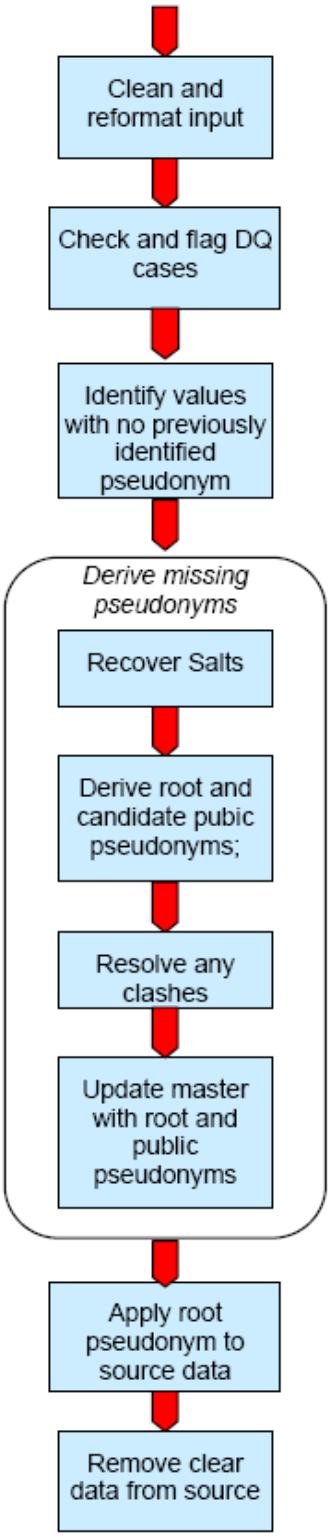
Acronym	Meaning	Notes
CDS	Commissioning Data Set	A nationally agreed list of data items that must be submitted to SUS (Secondary Uses System) for activity and PbR (Payment by Results) purposes for inpatient, outpatient and A&E attendances.
MDS	Minimum Data Set	A nationally agreed list of data items used to track activity and quality of services delivered by providers.
SUS	Secondary Uses System	A national data warehouse to which CDS datasets are submitted by providers for commissioners to access.
IAPT	Increasing Access to Psychological Therapies	An initiative to increase the availability of low and high intensity psychological help for people with anxiety and depression.
PbR	Payment by Results	A mechanism used by provider and commissioners to agree and pay based on tariff for inpatient, outpatient and A&E attendances.
PID	Patient Identifiable Data	Any data items or items which could be used to identify a specific person, e.g. NHS number, name, postcode, date of birth etc.

RBAC	Role Based Access Control	An IT security mechanism that allows access to systems and data to be controlled based on a person's job role.
SIRO	Senior Information Risk Officer	Nominated organisation official with overall responsibility for assessing information risks and ensuring that sufficient safeguarding are put in place to minimise or mitigate any risks identified.
SQL	Structured Query Language	A computer language developed specifically for manipulating and querying information held in a database.
PAS	Patient Administration System	A database system used by a service to record, monitor and track patients and patient appointments. These systems can also include clinical data recording facilities too.
ACL	Access Control List	A list of people or job roles who have permission to access an IT system and what level of access has been granted.
PIP	Pseudonymisation Implementation Project	Nationally mandated project for NHS organisations to implement pseudonymisation processes into their business processes where patient information is used for secondary purposes.

Appendix 1: De-identification methods to apply to sensitive data

Data item	How to de-identify	
	For Internal Use	**For External Use (via extract)**
Name	Do not display	Do not supply
Address	Do not display	Do not supply
Date of birth	Replace by age in years. For neo-nates use months instead.	Replace by age or age band in years. For neo-nates use months or banding instead.
Postcode	Postcode sector and/or derivations. Use Lower Super Output Area in preference to Output Area as the latter can be as small as 40 households.	
NHS Number	Pseudonymised or do not display	Pseudonymised / anonymised for one-off extracts or pseudonymised with consistent values (different values for different purposes for the same user) if for repeated extracts
Ethnic category	Identifiable if relevant to report, otherwise do not display	Do not provide unless relevant to the purpose of the analysis
Local patient identifier	Pseudonymised or do not display	
Hospital Spell Number	Pseudonymised or do not display	
Patient Pathway identifier	Pseudonymised or do not display	Pseudonymised
SUS Spell ID	Identifiable – but do not display if there is a potential to reveal confidential data through linkage	Pseudonymised
Unique Booking Reference Number	Pseudonymised or do not display	Pseudonymised
Social Services Client Identifier	Pseudonymised or do not display	
Date of death	Do not display unless relevant to the report in which case truncate to month and year.	Truncate to month and year

Appendix 2: The Pseudonymisation Process



Rules for applying pseudonymisation techniques are:

1. Each field has a different base for its pseudonym – e.g. with encryption, say key 1 for NHS Number, key 2 for date of birth; so that it must not be possible to deduce values of one field from another if the pseudonym is compromised.
2. Pseudonyms to be used in place of NHS Numbers and other fields that are to be used by NHS staff must be of reasonable length and formatted on output to ensure readability. Consideration also needs to be given to the impact on existing systems both in terms of the maintenance of internal values and the formatting of reports. For example, in order to replace NHS Numbers in existing report formats, then the output pseudonym should generally be of the same field length, (i.e. 10 characters but should not be only digits to avoid confusion with genuine NHS Numbers).
3. Pseudonyms generated from a hash must be seeded.
4. Pseudonymisation should be undertaken prior to user access and not 'on the fly', that is undertaking the generation of pseudonyms whilst processing and displaying data. This is because of the chances of error leading to inadvertent display of identifiable data. This method may be acceptable in producing extract files, but only if checks are made the output prior to dispatch to the user.
5. Pseudonyms for external use must be generated to give different pseudonym values (e.g. via use of a different hash seed) in order that internal pseudonyms are not compromised.
6. In the absence of explicit approvals to the contrary, data provided to external organisations should apply a distinct pseudonymisation to each data set generated for a specific purpose so that data cannot be linked across them. A consistent pseudonym may be required for a specific purpose, separate data sets are provided over a period of time and the recipient needs to link the data sets to create longitudinal records.
7. Display only the pseudonymised data items that are required, e.g. do not display pseudonymised date of birth if it is not relevant to a report
8. De-pseudonymisation requests for, and access to data in the clear, must be fully logged and approval by the appropriate authority documented.

9. Pseudonymised data must be treated in the same way as identifiable data in terms of security and access, as risks of re-identification do exist.
10. Pseudonymisation does not obviate the need to maintain the highest standards of security and confidentiality in local working and in system design and implementation.

Best Practice Recommendations:

1. Remove all white spaces (left, middle and right) before applying the pseudonymisation function to postcodes.
2. Add new entries to master lookup tables as they are found to ensure that the order within the table is randomised.
3. The Master Patient Index table should contain the patient's NHS Number, Name, Postcode, Sex and Date of Birth. While NHS Number must be used where available, the other information can be used for linkage when an NHS Number is absent.
4. Access to master lookup tables should be via SQL functions or procedures to prevent users viewing the whole table.
5. Access to patient identifiable information must be logged and should be done via the functions or procedures used to access it.
6. Add a unique index to the pseudonymisation field to ensure new entries to a master lookup table are unique.
7. Cryptographic hash functions (MD4, MD5, SHA-1 and SHA-2) should be used as they create a fixed length hash.
8. Encrypt the seed/salt values used in hash functions
9. Consider encrypting clear data in master tables for name, address, house number, postcode etc.

Appendix 3: Secondary use authorisation access request form

Secondary Use Authorisation Access Request Form

Staff Name:	
Service or Team:	
Data to be granted access to (Information Asset):	
Access Request Type:	Grant / Change / Remove *
Level of access required, e.g. to all data, or just parts:	
Reason access is required:	
Line Manager Name:	
Line Manager Signature:	
Date of Request:	

Person to approve request:	Information Asset Owner Caldicott Guardian Senior Information Risk Officer *
Name:	
Decision:	Access Approved / Not Approved *
Date of decision:	
Date Information Asset Register Updated:	

* Please delete as applicable

Appendix 4 - Equality Impact Assessment

Step 1 – Scoping; identify the policies aims	Answer
1. What are the main aims and objectives of the document?	To ensure that there is a fair and consistent approach to ensure that patient identifiable data is not used for secondary uses and that it has been appropriately de-identified when used for such purposes.
2. Who will be affected by it?	All Staff
3. What are the existing performance indicators/measures for this? What are the outcomes you want to achieve?	Not Applicable
4. What information do you already have on the equality impact of this document?	None
5. Are there demographic changes or trends locally to be considered?	No
6. What other information do you need?	None

Step 2 - Assessing the Impact; consider the data and research	Yes	No	Answer (Evidence)
1. Could the document unlawfully discriminate against any group?		X	The policy ensures all staff are treated in a consistent manner
2. Can any group benefit or be excluded?		X	The policy ensures all staff are treated in a consistent manner
3. Can any group be denied fair & equal access to or treatment as a result of this document?		X	The policy ensures all staff are treated in a consistent manner
4. Can this actively promote good relations with and between different groups?	X		Due to the consistency of approach everyone will be treated equally
5. Have you carried out any consultation internally/externally with relevant individual groups?		X	None required
6. Have you used a variety of different methods of consultation/involvement		X	None required
Mental Capacity Act implications			
7. Will this document require a decision to be made by or about a service user? (Refer to the Mental Capacity Act document for further information)		X	This document is aimed only at protecting patient data and not decisions made by them or about them.

If there is no negative impact – end the Impact Assessment here.

Step 3 - Recommendations and Action Plans	Answer
1. Is the impact low, medium or high?	
2. What action/modification needs to be taken to minimise or eliminate the negative impact?	
3. Are there likely to be different outcomes with any modifications? Explain these?	

Step 4- Implementation, Monitoring and Review	Answer
1. What are the implementation and monitoring arrangements, including timescales?	
2. Who within the Department/Team will be responsible for monitoring and regular review of the document?	

Step 5 - Publishing the Results	Answer
How will the results of this assessment be published and where? (It is essential that there is documented evidence of why decisions were made).	

****Retain a copy and also include as an appendix to the document****