Pseudonymisation Implementation Project (PIP)

Reference Paper 1

Guidance on Terminology

Version FV1.0 20 November 2009

Guidance on Terminology			
Programme	NPFIT	Document Record ID Key	
Sub-Prog / Project	Pseudonymisation Implementation Project (PIP)	NPFIT-FNT-TO-BPR-0023.01	
Prog. Director	J Thorp	Version	1.0
Owner	.	Status	Final
Author	J Fistein	Version Date	20 November 2009

Document Status:

This is a controlled document.
Whilst this document may be printed, the electronic version maintained in FileCM is the controlled copy. Any printed copies of the document are not controlled.

Related Documents:

These documents will provide additional information.

Ref	Doc Reference Number	Title	Version
1	NPFIT-FNT-TO-BPR-0022.01	PIP Implementation Guidance	FV1
	NPFIT-FNT-TO-BPR-0023.01	Reference Paper 1 - Terminology[1]	FV1
2	NPFIT-FNT-TO-BPR-0024.01	Reference Paper 2 – Business Processes and New Safe Havens[2]	FV1
3	NPFIT-FNT-TO-BPR-0025.01	Reference Paper 3 – De-identification[3]	FV1
4	TBA	Reference Paper 4 – Technical White Paper[4]	FV1
5	dh_4069254	NHS Code of Practice on Confidentiality[5]	
6	NA	PIP Planning Template and Guidance[6]	

[1] **http://www.connectingforhealth.nhs.uk/systemsandservices/sus/delivery/pseudo**
[2] **http://www.connectingforhealth.nhs.uk/systemsandservices/sus/delivery/pseudo**
[3] **http://www.connectingforhealth.nhs.uk/systemsandservices/sus/delivery/pseudo**
[4] **http://www.connectingforhealth.nhs.uk/systemsandservices/sus/delivery/pseudo**
[5] [5]

www.dh.gov.uk/en/Managingyourorganisation/Informationpolicy/Patientconfidentialityandcaldicottguardians/DH_4100550
[6] **http://www.connectingforhealth.nhs.uk/systemsandservices/sus/delivery/pseudo**

Contents

1 Introduction and Context

1.1 About this document

1.1.1 This document is a deliverable of the Pseudonymisation Implementation Project, a project delivered as part of the NHS CFH Secondary Uses Service (SUS), itself part of the NHS Information and Reporting Service (NIRS) Programme. It aims to define the terminology that should be use to describe different uses and types of information in the NHS.

1.2 Executive summary

1.2.1 This document proposes that the terms “primary” and “secondary” uses should be mapped clearly to the definitions in the *Confidentiality: NHS Code of Practice* **Ref 1** It proposes that the term “healthcare medical purposes” should be used preferentially for “primary uses” and the term “non-healthcare medical purposes” should be used preferentially for “secondary uses”. These terms are consistent with NHS guidance and the law. The document gives examples of particular activities that fit into each category.

1.2.2 The document also proposes terms for particular transformations of data that can be used to reduce the likelihood that individuals can be inferred from that data. It distinguishes these technical processes from their potential effects, which must be assessed in the particular context when data is being processed[7].

1.2.3 The term “effective anonymisation” is proposed for situations where data has been transformed such that there is no reasonable chance that individual identities could be inferred from the data in the circumstances that apply[8].

1.3 Context

1.3.1 Healthcare professionals create and maintain medical records to deliver safe and effective care of their patients. It is generally recognised that it is acceptable and legal for medical records to be used for such purposes in a form that can identify the patient. This is because it is presumed that using any other type of information for care might compromise the care of the individual concerned.

1.3.2 The information in medical records is additionally used for a variety of purposes outside the care setting. The NHS Care Record Development Board[9] (now part of the National Information Governance Board) has recognised the range and value of these uses, which include:

- Improving the quality of local clinical care, for example through the audit of clinical practice.
- Protecting the health of the public through surveillance and immediate response to infectious disease and other environmental threats to health,

[7] Data processing is a broad term that includes (but is not limited to) storage, transmission, display, transformation, etc.
[8] For example, from data that is transmitted in a file by the user of that file, from the online display of data, etc.
[9] E.g. The NHS Care Record Guarantee and the CRDB Report on the Secondary Uses of Patient Information.

monitoring adverse effects of therapeutic interventions, and informing and evaluating screening.

- Government policies and other initiatives for improving the management of the health system, for example by supporting the more efficient commissioning of services and to ensure that providers are reimbursed for the care services they provide.
- Identifying patients who interact with multiple parts of the health system in order to monitor equity of access and provision.
- Ensuring that health policy is evidence-based through carrying out empirical research.
- Providing better information to the general public about healthy lifestyles.
- [Indirectly] improving the quality and safety of care or reducing the impact of new risks to population health through for instance, research by the patient's clinical team; research by others using data collected by the care team but involving no contact with the team's patients; or research which requires further contact with patients or former patients.

1.3.3 Current law and guidance (such as the NHS Code of Practice on *Confidentiality*) recognises that such uses are permissible, but advises that "non identifiable" or "anonymised" patient information should be used for some of these uses. This position has caused confusion for NHS staff who are unsure about which specific purposes should be performed using "anonymised" data and about what should be done to data to ensure it is "anonymised". The latter issue is further complicated by the fact that data users feel that transforming data into anonymised forms reduces the utility of that data, and increases the likelihood that erroneous conclusions might be drawn from it.

1.3.4 This document aims to describe the techniques that can be used to transform patient information into forms which reduce the likelihood of disclosing patient identities, and when such techniques should be used by exploring and providing definitions for:

- Different contexts for data use.
- The technical processes that can be used to transform data into "less identifiable" or "anonymised" forms.
- The effects of such transformations on the legal status of the data.

2 Defining purposes of use of NHS data

2.1 Introduction

2.1.1 NHS data is used to deliver care, to support care indirectly and for other NHS purposes. Data use and sharing within the care setting is distinguished from other purposes in policy and law. However, this is currently done inconsistently, and using a variety of overlapping terms.

2.1.2 This document uses the terminology in the NHS Code of Practice on *Confidentiality* and attempts to map the terms "primary" and "secondary" uses to the terms in the Code of Practice.

NHS Guidance: Healthcare and Medical Purposes

2.1.3 Confidentiality: the NHS Code of Practice distinguishes between:

- "Healthcare purposes" i.e. those which "directly contribute to the diagnosis, care and treatment of an individual and the audit/assurance of the quality of the healthcare provided", and
- "Medical purposes" which are broader, and include "healthcare purposes" plus "preventative medicine, medical research, financial audit and management of health [and social] care services"[10].

2.1.4 It is proposed that this terminology should be followed, but the relationship between "medical" and "healthcare" purposes should be made explicit. The preferred terms should therefore be: "**healthcare medical purposes**" and "**non-healthcare medical purposes**". This reflects the fact that "healthcare purposes" are a type of "medical purposes". It is consistent with definitions of "medical purposes" in the Data Protection Act 1998 and the Health Service Act 2006. Examples of purposes that fit into each of these categories are given below.

"Primary" and "secondary" uses

2.1.5 Ideally the terms "primary" use and "secondary" use should not be used. These terms have no basis in law, and may cause confusion with data users as there is some argument about which particular purposes should be included in each of the categories. For example, audit has been classified as a secondary use by some and a primary use by others. Additionally, many professionals who use data for so-called "secondary" purposes object to the term, as it implies that these activities are of secondary importance, rather than, as they claim, essential to providing "the UK with universal, effective healthcare"[11].

2.1.6 However it is recognised that the terms "primary" and "secondary" uses are used widely in the NHS, for example in the names of legacy systems and

[10] p. 6. Although the Code of Practice focuses on Medical Purposes, it is recognised that the definitions in subsequent legislation also refer to social care data, as well as healthcare data.
[11] Academy of Medical Sciences, Personal data for public good: using health information for research, p. 29.

programmes. It should be made clear that “primary” uses equate to “healthcare medical purposes” and “secondary” uses equate to “non-healthcare medical purposes”. This position is summarised in Table 1 below.

Table 1: “Healthcare Medical Purposes” (“Primary uses”) and “Non-healthcare Medical Purposes” (“Secondary uses”)

Medical purposes	Healthcare medical purposes (aka “primary uses”	Uses which “directly contribute to the diagnosis, care and treatment of an individual”; or “the audit/assurance of the quality of the healthcare provided” **Person identifiable data can be used, but only the minimum amount of data should be used, and appropriate safeguards should be in place (see text)**
	Non-healthcare medical purposes (aka “secondary uses”)	Preventative medicine, medical research, financial audit and the management of health [and social] care services **Generally “effectively anonymised” data should be used, unless consent has been gained from the patient or there are special circumstances, such as an overriding public interest, or a route such as via Section 251 of the NHS Act 2006 or the Health Service (Control of Patient Information) Regulations 2002. However, current constraints on data quality reduces the ability to carry out such activities using effectively anonymised data, with the consequence that central NHS policy objectives cannot be realised. In the interim period therefore, where data and business processes are being refined in order to enable the use of effectively anonymised data, it may be necessary to use person identifiable data temporarily. However the amount of person identifiable data used should be minimised, and appropriate safeguards should be in place. (See text)**
Non-medical purposes	Court reports Police reports Etc. **Out of scope of the Pseudonymisation Implementation Project – See the NHS Code of Practice on *Confidentiality* for guidance**	

2.2 Healthcare medical purposes (Primary uses)

2.2.1 Healthcare medical purposes include two types of use:

- those which "directly contribute to the diagnosis, care and treatment of an individual"; and
- "the audit/assurance of the quality of the healthcare provided"

Using information to directly contribute to the diagnosis, care and treatment of an individual

2.2.2 Communicating with other colleagues in order to support the provision of diagnosis, advice and treatment of a patient is generally regarded as so closely bound up with the practice of medicine as to be indistinguishable from it. "Most people understand and accept the information must be shared within the healthcare team in order to provide the care"[12]. It has additionally been suggested that "a patient who consults a doctor impliedly consents to the doctor disclosing such information about the patient to other appropriately skilled staff…as may be necessary" to enable the doctor to decide how best perform these activities[13]. It is unlikely that non-person identifiable data (e.g. "pseudonymised" or "effectively anonymised" information) could be used to support many of these activities, and the use of such information may compromise the safety and quality of care.

2.2.3 Given the above position, it is generally accepted that information may be transmitted in person identifiable form for these activities, however this should be done:

- On a "need to know" basis;
- Within a secure system (technical and organisational);
- Transmitting the minimum amount of information required to provide safe care, and using techniques to reduce the likelihood that identities can be inferred from the data as long as this does not compromise the quality of care.

2.2.4 Activities in this category include:

- Face-to-face clinical interactions with the patient;
- Information required as part of a referral or treatment process (for example, use of a Patient Administration System, writing clinical letters by a clinician of a medical secretary);
- Activities directly supporting care such as management of a particular patient on a ward;
- Managing appointments for care.

[12] General Medical Council (UK), Confidentiality: Protecting and providing information, para.10.
[13] Toulson and Phipps, *Confidentiality* (2nd Edition), Sweet & Maxwell, 2006, 11-015.

Audit and the assurance of the quality of care

2.2.5 Health professionals have obligations to participate in clinical audit[14] as this is "essential to the provision of good care" [15] although the GMC states that patients should be informed of this use of information relating to them as far as possible[16]. It is permissible to use person identifiable information for these activities. However, where audits can be effectively performed using non-person identifiable data, this should be used wherever possible.

2.2.6 Activities classed as clinical audit may include:

- Auditing the quality of care within an organisation;
- Auditing the quality of care within a care pathway provided across several organisations;
- Monitoring the safety of interventions.

2.2.7 Information may be used in person identifiable form for these activities, however this should be done:

- On a "need to know" basis;
- Within a secure system (technical and organisational);
- Transmitting the minimum amount of information required to provide effective audit.

[14] General Medical Council, *Good Medical Practice,* para. 14.
[15] General Medical Council (UK), Confidentiality: Protecting and providing information, para.13.
[16] Ibid.

2.3 Non-healthcare medical purposes (Secondary uses)

2.3.1 These include:

- preventative medicine;
- medical research; and
- financial audit and the management of health [and social] care services

2.3.2 There has been an assumption that these purposes have always been possible using "effectively anonymised information" However, it has been argued that information that links to individuals (although not necessarily by name) is required for these purposes for several reasons[17]:

- To assess or avoid double counting of the same individual, which would bias research results or skew financial reports.
- To enable longitudinal research, which follows the health outcomes of individuals over extended periods of time.
- For linkage between data sets which depends on the reliable identification of individuals in different data sets.
- For validation of the quality of datasets i.e. to ensure that datasets are consistent and accurate, usually by cross-checking data relating to individuals from different sources.
- Where the identifiers are themselves required for the non direct care purpose, or where removal would render the work useless, for example where Postcodes are used to give an indication of deprivation status.
- When data cannot be anonymised, for example where data exist in paper form and cannot be examined independently of their associated identifiers.

2.3.3 The remainder of this section examines non-healthcare medical purposes in more detail.

Preventative medicine

2.3.4 Preventative medicine includes a variety of activities that aim to safeguard and improve public health, including public health monitoring and drug safety monitoring (in particular, including activities aimed at improving the quality of clinical data reporting). Many activities could be possible using "effectively anonymised" information, although some tasks (such as where it is necessary to follow-up individual patients) depend on the use of person identifiable information. The Health Service (Control of Patient Information) Regulations 2002 specifically permits processing using person identifiable information for some purposes – for example to link data sets together, for data quality purposes and to make the data less identifiable (for later use or reporting). Where data is used in person identifiable form, the same principles as above should apply i.e. information should be used:

[17] Academy of Medical Sciences, Personal data for public good: using health information for research, pp. 46-47.

- On a “need to know” basis;
- Within a secure system (technical and organisational);
- Transmitting the minimum amount of information required.

Research uses

2.3.5 The Health Service (Control of Patient Information) Regulations 2002 allow processing for research purposes regardless of any obligation of consent[18] provided this has been approved by the Secretary of State and a research ethics committee. In the absence of such approvals, data should only be used in an effectively anonymised form for research purposes.

Managing the health service

2.3.6 Legally, general practitioners and hospitals have to maintain computerised and other records in order to be able to manage their service to operate efficiently and properly. It has been argued that the keeping of such records does not infringe any duty of confidence to the patient, as it directly services the patient’s present and future treatment[19]. It has also generally been accepted that centralised, computerised record-keeping will be legal (even they hold person identifiable information), as long as appropriate security controls are in place to prevent unauthorised access.

2.3.7 Records are used more generally to manage the health service, for example to support central governmental policy initiatives such as:

- Ensuring that providers of care are properly reimbursed for the care they have provided – e.g. Payment by Results
- Reducing waiting times – the Referral to Treatment Initiative
- World class commissioning

2.3.8 The position on the use of confidential patient information[20] for these purposes is that their use may only be legal with the creation of new Regulations under s.251 of the NHS Act 2006. It is recognised, however, that many of these purposes are either mandatory or have significant financial implications for NHS organisations. Pragmatically, therefore, patient records should be used in an “effectively anonymised” form wherever possible and organisations should proactively seek ways of minimising the use of personal data.

2.3.9 NHS business processes will need to be modified or changed wherever possible to ensure that the transmission and handling of person identifiable data is minimised. For example, although it may not be possible currently to use effectively anonymised data for some provider-commissioner

[18] Section 5.
[19] Toulson and Phipps, para. 11-035.
[20] This may also apply to social care records that relate to service users and to confidential information that is not about patients.

conversations (because there is no way to ensure that the same patients are being discussed), this should not be taken as a warrant to continue with person identifiable flows indefinitely. Business process changes (for example, using consistent pseudonymisation from source systems) should enable conversations to be possible using effectively anonymised information.

2.3.10 In the 2009 guidance on Confidentiality published by the GMC it is made clear that where information is disclosed for non-healthcare medical purposes that express consent should be gained, information should be anonymised or s251 permission obtained. Doctors are advised to draw attention to any system that prevents them from following this guidance and recommend change. Until changes are made patients need to be informed about disclosures and what they should do if they object and if at all possible their objections should be acted upon. When information is disclosed, doctors should ensure that recipients are bound by a duty of confidentiality not to disclose it further.

3 Techniques for reducing the likelihood that patient identities can be inferred from data

3.1 Background

3.1.1 There are several technical processes that can be used to make it less likely that individuals can be identified from data, and these can be used singly or together at different points in the data lifecycle[21]. The choice and use of techniques has a critical impact not only on their effectiveness but also on the potential utility of the data.

3.1.2 A variety of terms has been used in the past to describe the technical means to turn data that can be traced to an individual into data that cannot. These terms have been found to be confusing (meaning different things to different people) and sometimes contradictory by many professionals who have to use and manipulate data routinely[22].

3.2 Anonymisation – a term that has no technical meaning

3.2.1 It is suggested that the term "anonymisation" should not be used in any technical communications. The more specific terms given in Table 2 should be used as these more accurately describe the process that is being used to transform the data[23].

Separation of techniques from effect

3.2.2 One source of confusion is the interchangeable use of terms to describe particular data transformations, the effects of such transformations on the "identifiability" of the data, and the legal status of transformed data sets.

3.2.3 It is important to separate out the effect of any process from the process itself. The simple use of a technical process to transform data into a "less identifiable" form might not be enough to discharge a data controller's obligations to protect e.g. patient confidentiality, as the "disclosiveness" of the information is determined by the knowledge of the recipient as much as by the mechanism used. For example, although a dataset has been pseudonymised, it may still be possible to infer the identities of particular individuals from that data due to the presence of rare diagnosis codes. Such dangers increase as more data items that are linked and associated together, where it becomes more likely that a data recipient could infer the identity of a patient even when obvious identifiers have been removed.

---

[21] Lowrance W., Learning from Experience: Privacy and the secondary use of data. Nuffield, London, 2002, p. 18.

[22] E.g. Academy of Medical Sciences, Personal data for public good: using health information for research, 2006, p. 48.

[23] The term "anonymisation" has been used variously:

- In a **non-technica**l way to mean, "Any tool in reducing the risk of harm from inadvertent disclosure", sometimes qualified as "strong", "weak" or "partial" anonymisation, depending on the degree of effectiveness in achieving this aim;
- To describe any **technical** process to make it less likely that an individual could be identified from a data sets (up to and including creating totally anonymous data); or
- Specifically the process of removing person identifiers from datasets. This latter confusion is particularly dangerous because (as described in this document) the simple removal of identifiers is seldom enough to render datasets truly anonymous and is therefore "not a sufficient strategy for protection against a deliberate attempt to breach confidentiality".

"Effectively anonymised" data

3.2.4 The aim of the technical processes described in this document is to reduce the chances that an individual's identity could be inferred from data by the recipient. The effectiveness of the particular techniques chosen must be assessed by the data sender given the context of the particular data transmission in question.

3.2.5 Where there is no reasonable chance that the recipient could infer identities from that data, the data can be termed "**effectively anonymised**". It is recognised that such data that falls short of the definition of "completely anonymous" data, from which it would be *impossible* for *any* recipient to infer identities from the data, however "effectively anonymised" data would almost certainly neither be considered "personal data" nor "sensitive personal data" under the DPA 1998, nor "confidential patient information" under the NHS Act 2006.

3.2.6 For data to be "effectively anonymised" the recipient must be unable to infer the identity of individuals from that data without the application of effort or resource that it would be unreasonable to anticipate in the circumstances that apply. For example, data senders must consider all potential routes to identification of data when sending the data, and must adjust what they send accordingly.

There is no technique of "effective anonymisation"

3.2.7 It should be stressed that there is no simple **technical** process of "effective anonymisation". Rather, combinations of techniques applied to the data and controls on the recipient will need to be chosen as appropriate for the given circumstances to ensure the recipient cannot infer identities from the data. This overall process may be termed "effective anonymisation" but it must be stressed that this refers to the effect of any technical process and does not imply the use of any particular technique. Data that has not been effectively anonymised should be described as such i.e. "not effectively anonymised information".

3.3 Techniques to make data less likely to identify individuals

3.3.1 For the purposes of the Pseudonymisation Implementation Project, the following terms are used to describe the **technical** processes that can be used to transform data to make it less likely that individuals can be identified. The preferred generic term for these processes is a whole is **De-identifying data** although it should be stressed that the processes **do not** create wholly "deidentified" data (this implies a level of anonymity that is potentially not justified) and so should not be referred to as "deidentification".

3.3.2 Several terms have been used for these technical processes in different communities and these are given where they are known.[24]

3.3.3 It must be stressed again that the use of any particular technique does not guarantee "anonymity" of the data, or any discharging of any obligations under the Data Protection Act, other law, or NHS policy. Such techniques may reduce the risk of identification, but there must be a separate assessment of whether the reduced risk is acceptable in particular circumstances and whether the technical process constitutes "effective anonymisation" if this is required.

3.3.4 It should also be noted that even where it is legally permissible to use person identifiable information, its use should be minimised, for example by only sending a relevant subset of the information or by the use of one or more of the techniques described.

[24] Many of these other terms are vague, potentially misleading or conflate several techniques and their effects. It is therefore difficult to map these legacy terms accurately to the proposed preferred terms.

Table 2: Technical processes used to de-identify data[25]

Preferred technical term	**Technical description**	**Terms used elsewhere** (and reasons why these should not be used where these are compelling)	**Example**
Identifiable information	(Usually) a (source) dataset that includes person identifiers that will ordinarily and simply identify a person. It may be "personal data" under the DPA or "sensitive personal data" under the DPA if it relates to the health of the patient. It may also be "confidential patient information" under the NHS Act 2006. Disclosure of such information may breach a common law duty of confidence and / or may infringe individuals' rights to privacy. When the persons in question are patients, the data may be called "patient identifiable data".	**Personal information / data** - this term should not be used as it confuses techniques and effect. The term also has a specific meaning in the Data Protection Act. Other types of data in this table may be considered personal data. **Nominative information** – it is not only names that make data person identifiable. **Clear** – although this term is used in the NHS, research users find the term confusing and misleading.	Data that that contains one or more person identifiers. Some of these are **direct** or **strong** identifiers such as: • Name and address • Date of Birth (DoB) • Postcode • NHS Number, Local Patient Identifier, etc. i.e. symbolic identifiers that relate to a particular individual that can be simply decoded Others may be **indirect** or **weak** identifiers found in the clinical "payload" data such as: • Dates of clinical encounter • Ethnicity • Rare diagnosis

[25] The term "privacy enhancing technology" has been used for such techniques, but this term fails to include protection relating to other statutory and common law principles.

Preferred technical term	**Technical description**	**Terms used elsewhere** (and reasons why these should not be used where these are compelling)	**Example**
Stripping out (or not displaying) person identifiers data to create a **data set in which person identifiers are not present**	Removing person identifiers from the data. May be partial (where some identifiers are removed) or complete. .	**Non-identifiable data / Unidentifiable data** – these terms are misleading, as the absence of person identifiers does not necessarily mean patient identities could not be determined (e.g. where there is the presence of a rare diagnosis) **Anonymising data / anonymous / unlinked anonymised data** – as above; the data is not necessarily "anonymous"	A dataset that comprises records which only "payload" data, without any person identifiers.

Preferred technical term	**Technical description**	**Terms used elsewhere** (and reasons why these should not be used where these are compelling)	**Example**
Pseudonymising data to create a **Pseudonymised** data set	The technical process of replacing person identifiers in a dataset with other values (pseudonyms) available to the data user, from which the identities of individuals cannot be intrinsically inferred, for example 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.).	**Coding / Key-coding** – "coding" could relate to the coding e.g. of diagnoses codes. **Reversibly De-identifying** – De-identifying data has been given the more specific and accurate definition above. Pseudonymisation may be reversible or irreversible. Performing linked anonymisation / Pseudo-anonymisation – as above, the data is not necessarily anonymised, nor is there necessarily any linkage. **Masking –** too ambiguous a term; could imply masking person identifiers e.g. with blanks not pseudonyms. **Encrypting** – this refers to the means of transmission of data. Person identifiable data or pseudonymised data can be encrypted when transmitted.	Data that contains • A pseudonym or collection of pseudonyms (for example, one for name, one for DoB, etc.) Plus non-identifying "payload" data.

Preferred technical term	**Technical description**	**Terms used elsewhere** (and reasons why these should not be used where these are compelling)	**Example**
Aggregating data to create an **Aggregated** data set	Pooling data such that totals in categories are displayed not individual values. Must also **inference control techniques,** such as small number suppression, Barnardisation[26], etc. which prevent the inference of individual identities from small cell values.	**Banding** – this term is generally used to mean a type of derivation as descried immediately below, therefore should not be used to mean "aggregating data" as defined here. **Grouping**	Data is displayed as totals, so no individual data is shown. Small numbers in totals are suppressed.
Using derivations to create a dataset that contains **derived data items**. **Banding** data (to create a **banded** data set)	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, DoB replaced by an age range) this is **Banding.**	**Masking –** see above. **Displaying partial data items** – too non-specific **Data blurring**	Data that contains individual data items with: • Area or ward in place of address • Age in place of DoB • Partial postcode or area in place of postcode • Etc.

[26] A method of disclosure or inference control for tables of counts that involves randomly adding or subtracting 1 from some cells in the table.

Preferred technical term	**Technical description**	**Terms used elsewhere** (and reasons why these should not be used where these are compelling)	**Example**
Using synthetic data techniques to create a **synthetic data set**	Mixing up the elements of a dataset so that all of the totals and values of the set are preserved but are not related to any particular individual. Such data would allow overall totals and frequencies to be calculated accurately. This type of data may be useful for system testing.	**Perturbation**	Data that contains • Name and address • DoB • Postcode • NHS Number, Local Patient Identifier, etc. i.e. symbolic identifiers that relate to a particular individual that can be simply decoded Plus "payload" data – e.g. clinical diagnosis, treatments, etc. **However,** the data items are shuffled such that the data items in a particular row do not relate to the same individual.

Preferred technical term	**Technical description**	**Terms used elsewhere** (and reasons why these should not be used where these are compelling)	**Example**
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.	**Geographic sequestration** – only refers to a particular type of quarantining.	Usually used in combination with one or more other techniques, so data appear as above.

Reversibility of techniques

3.3.5 Some of the techniques described above are **technically irreversible** i.e. there is no technical means to get back to the original identity of the patient from the data supplied. For example:

- If data is supplied to a recipient without any person identifiers it will be technically impossible to then confirm that the data relates to any given individual. If further data is supplied to the recipient it will be impossible for the recipient to link together data that relates to any particular individual.
- Where data is provided to recipients in aggregated form, it is generally impossible to reverse the aggregation process to get back to individuals.

3.3.6 Even though it may be technically possible to reverse some other techniques above and therefore to get back to information that can identify individuals there should be procedures in place to ensure that any reversal (if allowed at all) is appropriate in terms of the business use of the data, IG policies and law. It is the responsibility of the sender to ensure that appropriate measures are in place or otherwise they would be considered reckless for failing to effectively mitigate the risks of identification of individuals from the data.

3.4 Techniques to reduce the risk that the recipient is able to infer identities from data

3.4.1 As described above, where “effectively anonymised” data should be used for a particular purpose, the data sender must make sure there is no reasonable chance that the recipient could infer identities from that data. The table above describes techniques to transform data into forms where it may be less likely that identities can be inferred from data, however, the sender may also impose conditions on the recipient to ensure that the recipient handles the data appropriately. For example:

- Assessing and mitigating for the knowledge of the recipient – thus some data might be “effectively anonymous”, if the data is not person identifiable and there is a negligible chance that the recipient (has other knowledge or the means to analyse) the data to work out the identities of individuals.
- Assessing the nature of the recipient and the threat – for example the risk of finding the identity of a particular individual, of any individual, or about groups of individuals.
- Placing conditions on the recipient, e.g. confidentiality or privacy clauses in contracts, insisting on audit of use, and ensuring there are effective sanctions for any misuse (such as attempts to infer identities from the dataset).

INDEX OF TERMS