Information Standards Board for Health and Social Care			
Title	Anonymisation Standard for Publishing Health and Social Care Data Specification (Process Standard)		
Document ID	ISB1523 Amd 20/2010		
Sponsor	Phil Walker	**Status**	Final
Developer	Clare Sanderson & Malcolm Oswald	**Version**	1.0
Author	Malcolm Oswald	**Version Date**	21/02/2013

Anonymisation Standard for Publishing Health and Social Care Data Specification

Amendment History:

Version	Date	Amendment History
0.1	20/04/2012	First draft for comment by ISB. Previous drafts have been subject to considerable review (see front pages of: Drawing the line between identifying and non-identifying data).
0.2	30/05/2012	Amended specification and conformance criteria
0.3	27/06/2012	Amended following ISB Appraisal, NIGB review and Tech Office review
0.4	05/07/2012	Amended to make changes following minor comments from ISB reviewer
0.5	10/08/2012	Amended to reflect ISB Board comments, including merging the implementation guidance into this document.
0.6	26/09/2012	Amended to reflect comments from ISB Domain Leads and others
0.7	30/10/2012	Version used for phase 2 testing having been amended to reflect feedback from phase 1 testing
0.8	22/11/2012	Changes made to reflect the results of phase 2 testing
0.9	20/12/2012	Changes made to reflect feedback from ISB appraisers
0.91	21/01/2013	Changes made to reflect further comments from ISB
1.0	21/02/2013	Publication copy

Approvals and Reviews:

Name	Organisation & Role	Version	Review/ Approval
Information Standards Managers	Information Standards Management Service	V0.1 V0.2 V0.3	R
Stakeholder Groups	Health and Social Care	V0.2	R
Appraisers	Information Standards Board	V0.2	R
Phil Walker	Department of Health – Sponsor – Head of IG Policy	V0.1 V0.2 V0.3	R
Clare Sanderson	The Information Centre for Health and Social Care – Lead Developer – Executive Director of IG	V0.1 V0.2 V0.3	R
Phil Walker	Department of Health – Sponsor – Head of IG Policy	V0.9	A
Clare Sanderson	The Information Centre for Health and Social Care – Lead Developer – Executive Director of IG	V0.9	A
Technology Office	NHS Connecting for Health	Draft Submission	A

Karen Thomson (NIGB)	National Information Governance Board - IG Lead	Draft Submission	R
Information Standards Board	NHS Connecting for Health	Draft Submission	A
Information Centre Testers and other reviewers	Health and Social Care Information Centre	v.05, v.06	R
Information Standards Board Domain Lead and Appraisers	NHS Connecting for Health	V0.7 and V0.8	R
Information Standards Board	NHS Connecting for Health	V0.9	A
Information Standards Board	NHS Connecting for Health	V1.0	A

Glossary of Terms

Term	Definition
Aggregate data	Data derived from records about more than one person, and expressed in summary form, such as statistical tables.
Anonymisation	Any processing that minimises the likelihood that a data set will identify individuals. A wide variety of anonymisation techniques can be used; some examples of such processing are explained in this specification. Also commonly referred to as “de-identification”.
Caldicott Guardian	A senior person responsible for protecting the confidentiality of patient and service user information and enabling appropriate information sharing. Caldicott Guardians were mandated for NHS organisations by Health Service Circular HSC1999/012 and later for social care by Local Authority Circular LAC 2002/2. General practices are required by regulations to have a confidentiality lead[1]. Note that a Caldicott Guardian is an individual, whereas a data controller is a “legal person” (invariably an organisation such as a general medical practice or foundation trust).
Cell	An entry in a table of aggregate data.
Confidential information	Information to which a common law duty of confidence applies.
Data controller	A person[2] who (either alone or jointly or in common with other persons) determines the purposes for which and the manner in which any personal data are, or are to be, processed[3].
Data	Data means information which – (a) is being processed by means of equipment operating automatically in response to instructions

[1] The definition provided in the glossary that forms part of the Information Governance Toolkit, available at: https://www.igt.connectingforhealth.nhs.uk/Resources/Glossary.pdf
[2] Note that this is a “legal person”, and in the context of health and social care, this will be a legal entity such as a local authority, NHS trust, or general practice rather than an individual working for such a body.
[3] The definition in the Data Protection Act 1998; see: http://www.legislation.gov.uk/ukpga/1998/29/part/I

	given for that purpose, (b) is recorded with the intention that it should be processed by means of such equipment, (c) is recorded as part of a relevant filing system or with the intention that it should form part of a relevant filing system, (d) does not fall within paragraph (a), (b) or (c) but forms part of an accessible record as defined by section 68, or (e) is recorded information held by a public authority and does not fall within any of paragraphs (a) to (d)[4].
Data processor	Any person (other than an employee of the data controller) who processes the data on behalf of the data controller[5].
Direct identifier	Name, address, widely-used unique person or record identifier (notably National Insurance Number, NHS Number, Hospital Number), telephone number, email address, and any other data item that on its own could uniquely identify the individual.
Disclose	To provide information to specific recipient(s).
Duty of confidence	A duty of confidence arises when one person discloses information to another (e.g. patient to clinician) in circumstances where it is reasonable to expect that the information will be held in confidence. It – a. is a legal obligation that is derived from case law; b. is a requirement established within professional codes of conduct; and c. must be included within NHS employment contracts as a specific requirement linked to disciplinary procedures[6].
Identifying data	The same meaning as personal data, but extended to apply to dead, as well as living, people.

[4] The definition in the Data Protection Act 1998; see: http://www.legislation.gov.uk/ukpga/1998/29/part/I
[5] The definition in the Data Protection Act 1998; see: http://www.legislation.gov.uk/ukpga/1998/29/part/I
[6] Definition taken from page 7 of NHS Code of Practice on Confidentiality, available at: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4069253

Indirect identifier	A data item (including postal code, gender, date of birth, event date or a derivative of one of these items) that when used in combination with other items could reveal the identity of a person. Also referred to as "quasi-identifier".
Individual-level data	Data that have not been aggregated[7] and that relate to an individual person, and/or to events about that person. The data may or may not reveal the identity of a person, and thus may or may not be identifying data. An example is a request for an investigation that accompanies a blood test, with NHS Number, date of birth, and details of the sample and tests required[8].
Information	See definition of "data". Within this document, the two terms are used synonymously.
k-anonymity	A criterion to ensure that there are at least k records in a data set that have the same quasi-identifier values. For example, if the quasi-identifiers are age and gender, then it will ensure that there are at least k records with 45-year old females[9]. Note that it is necessary to remove direct identifiers in order to satisfy k-anonymity.
Non-identifying data	Data that are not "identifying data" (see definition above). Non-identifying data are always also non-personal data.
Non-personal data	Data that are not "personal data". Non-personal data may still be identifying in relation to the deceased (see definition of "identifying data" and "personal data").

[7] Note though that a record about an individual may contain aggregate counts (such as a data item of "Number of hospital admissions for patient in 2011".
[8] In this particular example, the data would be identifying because of the presence of the NHS Number and date of birth.
[9] The definition provided at page 47 of 'Best Practice' Guidelines for Managing the Disclosure of De-Identified Health Information', available at: www.ehealthinformation.ca/documents/de-idguidelines.pdf

Personal data	Data which relate to a living individual who can be identified – (a) from those data, or (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller, and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual[10].
Pseudonymisation	A technique that replaces identifiers with a pseudonym[11] that uniquely identifies a person. In practice, pseudonymisation is typically combined with other anonymisation techniques.
Publish	To disseminate to the public[12]. Note that "disseminate" is sometimes given a meaning similar to that of "disclose" above, although its dictionary meaning used here is quite different: "to spread abroad" and "to disperse throughout"[13].
Public authority	A body defined under, and subject to, the Freedom of Information Act 2000[14]. It includes government departments, local authorities, the NHS, state schools and police forces. A non-public authority, such as a company, carrying out health and social care activities on behalf of, and under contract to, a public authority may be required to assist the public authority in satisfying requests under the Freedom of Information Act.
Quasi-identifier	See entry for "indirect identifier" above.

[10] The definition in the Data Protection Act 1998; see: http://www.legislation.gov.uk/ukpga/1998/29/part/I
[11] A pseudonym is a fictitious name or code.
[12] A definition provided by Merriam-Webster Dictionary at: http://www.merriam-webster.com/dictionary/publish. To "Disseminate" here means unrestricted distribution or availability, and "the public" includes any person or group of people..
[13] A definition provided by Merriam-Webster Dictionary at: http://www.merriam-webster.com/dictionary/publish
[14] For a full definition, see Schedule 1 of the Freedom of Information Act at: http://www.legislation.gov.uk/ukpga/2000/36/schedule/1?view=plain

Redact	To censor or obscure (part of a text) for legal or security purposes[15]. In the context of responding to Freedom of Information Act requests, redaction should be explicit and permanent, making clear that information has been withheld.
Re-identification	The process of discovering the identity of individuals from a data set by using additional relevant information.
Statistical disclosure control	Techniques for obscuring small numbers (e.g. less than "5") that appear in aggregate tables so as to prevent re-identification.

[15] The definition provided by Oxford Dictionaries at: http://oxforddictionaries.com/definition/redact

Contents

1 Overview

1.1 Summary

Standard	
Standard Number	ISB 1523
Standard Title	Anonymisation Standard for Publishing Health and Social Care Data
Description	The law pulls in two opposite directions. Human Rights and Data Protection legislation, along with our domestic common law duty to respect confidentiality, require us to protect information that could identify an individual. The Freedom of Information Act requires public authorities to release information about their activities, and this message is reinforced by the government's transparency agenda. Although the law makes a clear distinction between identifying and non-identifying data, where that line should be drawn may be far from clear in practice. This anonymisation standard for publishing health and social care data is needed in order to address these difficult issues. This process standard provides an agreed and standardised approach, grounded in the law, enabling organisations to: • distinguish between identifying and non-identifying information, and • deploy a standard approach and a set of standard tools to anonymise information to ensure that, as far as it is reasonably practicable to do so, information published does not identify individuals.
Applies to	The following persons or bodies must comply with this process standard: a. The Secretary of State (DH) b. The NHS Commissioning Board (once established) c. any public body which seeks to publish information relating to the provision of health services or of adult social care in England; d. any person commissioned (by a public body) to provide such health services or adult social care who seeks to publish information relating to them; e. any person registered as described in s20A of the Health & Social Care Act 2008. It relates to the activities of the above bodies when processing[16] health and social care data, whether held electronically or on paper, and includes any data for which they are responsible, including personal data processed on

[16] This is the broad concept of processing used in the Data Protection Act,that "means obtaining, recording or holding the information or data or carrying out any operation or set of operations on the information or data…"..

	their behalf by a data processor. There are certain specific requirements when publishing official statistics[17]. Whilst broadly consistent with this standard, there are some specific principles and procedures in the Code of Practice for Official Statistics to be followed, and statisticians and others publishing official statistics should refer to, and apply, that code of practice, and the supporting guidance published by the Government Statistical Service[18]. The Information Commissioner's Office (ICO) has confirmed that this specification is consistent with the ICO's Anonymisation Code of Practice.
Release	
Release Number	Amd 20/2010
Release Title	Initial Release
Description	
Implementation Completion Date	30 April 2013

1.2 Related Documents

This Specification should be read in conjunction with the following documents:

Title
Anonymisation standard for publishing health and social care data – supporting guidance

1.3 Related Standards

This Specification should be read in conjunction with the following standards:

Standard No.	Document Reference	Title	Web link
ISB 1512	Amd 159/2010	Information Governance Framework	http://www.isb.nhs.uk/documents/isb-1512/amd-159-2010/index_html
ISB 0086	Amd 05/2011	Information Governance Toolkit	http://www.isb.nhs.uk/documents/isb-0086/amd-05-2011/index_html/
ISB 1572	Amd 97/2010	Sensitive Data	http://www.isb.nhs.uk/documents/isb-1572/dscn-41-1998/

The Anonymisation Standard fits within the Information Governance Framework.

[17] 'Official statistics' are defined in Section 6 of the Statistics and Registration Service Act 2007.
[18] Available at: http://www.statisticsauthority.gov.uk/national-statistician/ns-reports--reviews-and-guidance/national-statistician-s-guidance/index.html

1.4 Contacts

Sponsor	
Name	Phil Walker Head of Information Governance Policy
Organisation	Department of Health
Email Address	phil.walker@dh.gsi.gov.uk
Developer	
Name	Clare Sanderson Executive Director of Information Governance
Organisation	NHS Health & Social Care Information Centre
Email Address	Clare.Sanderson@ic.nhs.uk or malcolm.oswald@nhs.net
Developer	
Name	Malcolm Oswald – Subject Matter Expert
Organisation	NHS Health & Social Care Information Centre
Email Address	Clare.Sanderson@ic.nhs.uk or malcolm.oswald@nhs.net
Implementation Manager	
Name	Dawn Foster Head of Information Governance
Organisation	Health & Social Care Information Centre
Email Address	enquiries@ic.nhs.uk
Maintenance Manager	
Name	Michael Goodson Information Governance Manager
Organisation	Health & Social Care Information Centre
Email Address	enquiries@ic.nhs.uk

2 Specification

The key words MUST, SHOULD and MAY are standard terms used in the information standards methodology, and follow the meanings set out in RFC-2119:

- ***MUST*** - This word, or the terms "***REQUIRED***" or "***SHALL***", means that the definition is an absolute requirement of the specification.
- ***SHOULD*** - This word, or the adjective "***RECOMMENDED***", means that there may exist valid reasons in particular circumstances to ignore a particular item, but the full implications must be understood and carefully weighed before choosing a different course.
- ***MAY*** - This word, or the adjective "***OPTIONAL***", means that an item is truly optional."

2.1 Information Specification

#	Requirement
1	All Health and Social Care bodies choosing or obliged by law to publish (electronically or on paper) information/data relating to, or derived from, personal identifiable records MUST anonymise information so that information published does not identify individuals.

2.2 Conformance Criteria

This section describes the tests that can be measured to indicate that this process standard is being used correctly by an organisation (conformance criteria). These may be different depending upon the type of organisation, e.g. supplier, Trust, GP practice. The Information Governance Toolkit will be updated to reflect this standard and will be one means for measuring conformance.

#	Conformance Criteria
1	Health and Social Care bodies choosing or obliged by law to publish information/data relating to, or derived from, personal identifiable records MUST have regard to this process standard.
2	When publishing information after 1 April 2013, affected organisations MUST either: a) follow this standard; or b) follow alternative guidance of a similar standing.
3	If alternative guidance of a similar standing is used, affected organisations MUST record their reasons for choosing the alternative, and make their reasons available on request.
4	Whether this standard or alternative guidance is used, affected organisations MUST conduct, record, and make subsequently available on request, a risk assessment regarding the possibility that specific individuals might be identified from the published material either directly or indirectly through association of the published material with other information/data in or likely to be placed in the public domain.
5	Whether this standard or alternative guidance is used, affected organisations MUST record, carry out, and make subsequently available on request, an anonymisation plan, and SHOULD record their reasoning for choosing that plan. A spreadsheet for this purpose is provided and

	MAY be used.
6	Whether this standard or alternative guidance is used, affected organisations MUST, prior to publishing, confirm with the organisation's Caldicott Guardian or other responsible officer that the information to be published does not identify individuals, and this confirmation MUST be recorded and be available subsequently on request.
7	Where data previously published by the affected organisation are found to have led to confidential information about an individual being revealed, organisations SHOULD carry out an investigation into the incident and review their procedures for anonymising and publishing health and social care data. Any concerns, or suggested improvements, relating to this standard SHOULD be notified to the Health and Social Care Information Centre at: enquiries@ic.nhs.uk.
8	Organisations using the standard may wish to conduct a periodic audit to check the process is being followed and that appropriate judgements are being made by staff using the standard..

3 Implementation and Use

3.1 Guidance

Implementation Guidance is contained within sections 4-6.

3.2 Governance

Governance (e.g. the role of the Caldicott Guardian) forms part of the implementation guidance set out in *sections 4-6*.

3.3 Supporting information

It is recommended that the following document is read before applying this specification: *Supporting Guidance: Drawing the line between identifying and non-identifying data*.

4 Implementation guidance: standard anonymisation processes for health and social care organisations

4.1 Introduction

The purpose of this standard is to assist, and give confidence to, health and social care organisations transforming data that identify individuals into data that are not identifying and fit for publishing. Identifying information that reveals something confidential about a person must not be published. Transforming identifying data into non-identifying data protects personal privacy, and enables published information to be used for public benefit.

The anonymisation processes set out in this section are based as closely as possible on the relevant law described in the supporting guidance document "Drawing the line between identifying and non-identifying data". In particular, the standard is based on conclusions drawn directly from the concept of "personal data" in the Data Protection Act, namely that:

- If there is little or no possibility that a person will be identified from data released to others, then those data are not personal data, and as defined in the guidance, "non-identifying data"[19]; and
- Whether data are identifying depends not only on the intrinsic identifiability of the data, but on the context in which it is used, and specifically on the risk of additional information being used to reveal identity.

That guidance document also contains a glossary, statements on scope, and scenarios illustrating the application of the anonymisation standard. **Readers are advised to review the supporting guidance document on anonymisation before reading this document**.

Section 4 of this document contains a set of flowcharts, with supporting definitions, describing a standard process for publishing non-identifying data (see section 4.2). It draws on a sub-process entitled "Assess risk and specify data anonymisation" (see section 4.3).

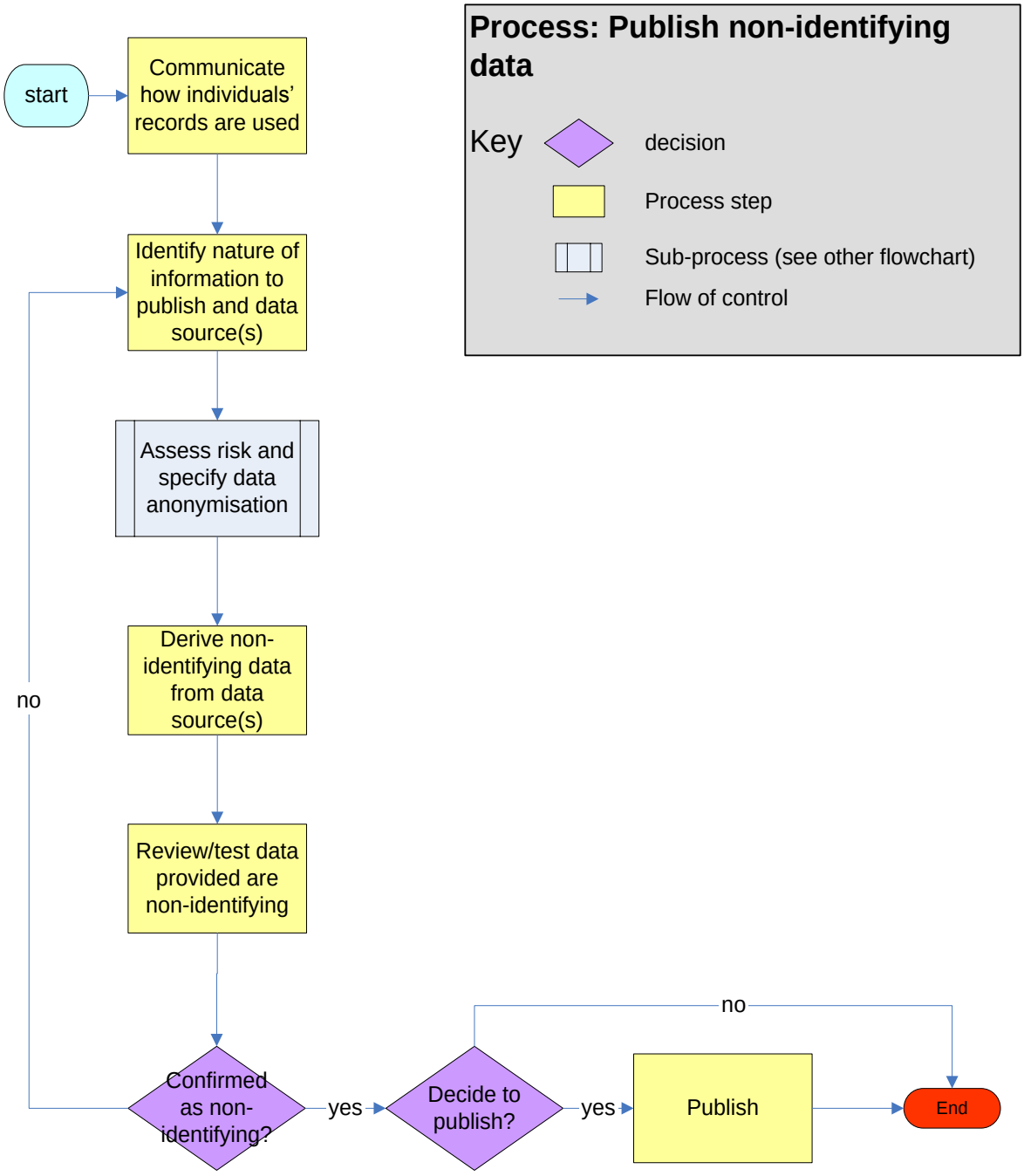
The processes and sub-process are each specified using a flowchart, followed by backing process and sub-process step descriptions.

The processes provide a standard approach to anonymisation when publishing, including a set of standard anonymisation plans. The standard anonymisation plans are specified in section 5, and the standard anonymisation techniques on which the plans are built, are specified in section 6.

As the context in which anonymisation is carried out, and the data that are used to derive non-identifying data, are so varied, one of the most important steps in the sub-process "Assess risk and specify data anonymisation" is "Refine anonymisation plan and specify anonymisation". This step recognises the importance of adapting the anonymisation plan to the specific circumstances of each case.

[19] Unlike "personal data", the concept of "identifying data" includes data about the deceased.

4.2 Process: Publish non-identifying data



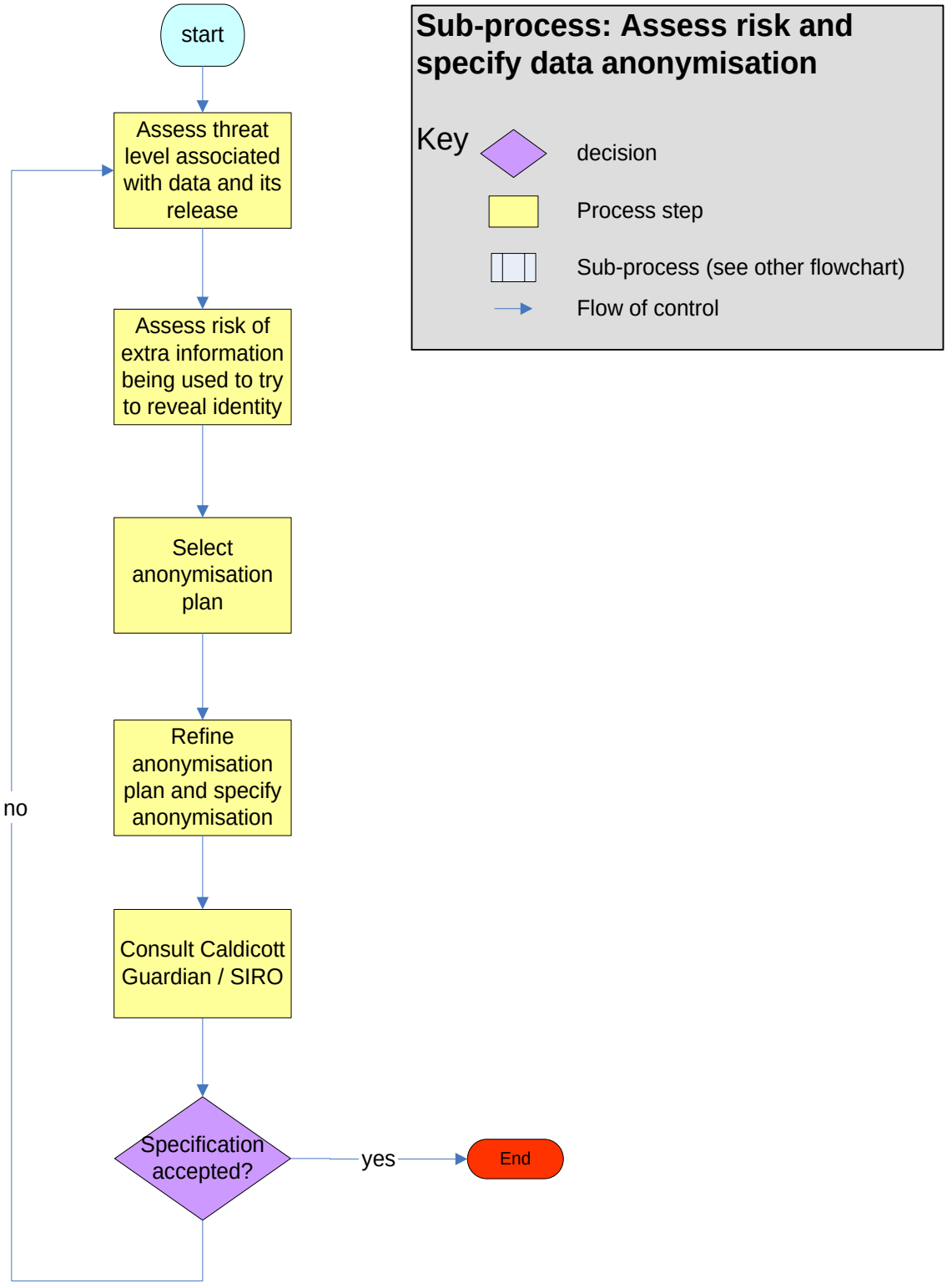
The table below describes the process steps that appear in the preceding flowchart.

Step name	Process step description
Communicate how individuals' records are used	**Provide information in an appropriate form to the public about how personal data from their records are being processed (including anonymisation processing).** This could be done for example, through the organisation's website, or with local posters. Information provided should be kept up-to-date, and should explain clearly the general nature and purpose of processing carried out by the organisation, but need not describe specific processing carried out to produce a particular publication. It should explain what action people can take if they object to the way their records are being used. This step is necessary to meet the fair processing requirement in the Data Protection Act. Furthermore, surveys have shown that the public wants to know how their records are used. Similar principles apply when publishing information derived from personal data about staff.
Identify nature of information to publish and data source(s)	**Consider the reasons or requirement to publish, what information ideally should be published in order to provide the greatest utility to the public and other potential users of the information, the structure of that information (e.g. if it should be broken down geographically or according to certain categories), and identify the data sources that would be required to produce it.**
Assess risk and specify data anonymisation	*See sub-process flowchart and descriptions to follow.*
Derive non-identifying data from data source(s)	**Apply the anonymisation specification agreed with the Caldicott Guardian[20] (or confidentiality lead) for the organisation and thereby create the non-identifying data from the data sources.**
Review/test data provided are non-identifying	**Review the data to be released, and ensure that:** **• the anonymisation steps specified (if any) have been carried out accurately;** **• the data being made available contains no information**

[20] Where the data relates to information other than patients, then the Senior Information Risk Owner should be consulted. See http://www.connectingforhealth.nhs.uk/systemsandservices/infogov/igap/igaf/roles.pdf

Step name	Process step description
	that is likely to reveal the identity of an individual once released; and **• record what testing has been done to ensure the data are non-identifying.** Where the potential impact of revealing identity would be significant (as for instance with sensitive information about abortions), or where there is a significant likelihood that the data to be released will be subject to attack by people seeking to reveal personal identity, consider carrying out testing on the data set to check it does not reveal identity, making use of commonly available demographic data (such as the electoral roll, phone directories) as well as any specific data sources relevant to the nature of the information being published. If in doubt, consult the Caldicott Guardian or other responsible officer.
Publish	**Make available the non-identifying data to the intended audience.** Once satisfied that the data to publish are non-identifying, that it is worth publishing (eg what can be published is still useful), and that there is authorisation to publish it, then publish the data. Publication may be in many different forms.,

4.3 Sub-process: Assess risk and specify data anonymisation



The table below describes the sub-process steps that appear in the preceding flowchart.

Step name	Process step description
Assess threat level associated with data and its release	**Assess the threat level as either "normal" or "high", and record your reasoning.** Amongst those that might be motivated to re-identify individuals from data sets that are intended to be non-identifying, certain data are likely to be of much greater interest than others. Much of the activity carried out in health and social care is relatively mundane and of limited interest to others, and thus associated with "normal" threat. Few people would seek to re-identify individuals from figures for admissions to nursing homes or hospital medical wards. In general, health and social care data should be considered "normal" threat. However, specific cases will be "high" threat. The likely sensitivity of the data can be an indicator of high threat in some cases. Information on termination of pregnancies in Leeds is particularly sensitive and of interest to certain people, and thus more likely to be a target for re-identification, than information about hip operations in Leeds. Information published by a hospital that might reveal the reason that a celebrity was admitted to hospital might have commercial value. Some people may be motivated by the status gained in penetrating certain targets that ought to be particularly secure, such as the medical records of high-security prisoners, or a national register of sex offenders. A man seeking his ex-wife and children may be particularly motivated to discover their address in a women's refuge in Sheffield. Thus certain people are going to be motivated to try to re-identify individuals within a data set if the nature of the data is of particular interest or value to them, or if successful re-identification offers a certain status. Motivation is the major determinant of the threat level associated with the data and its release.
Assess risk of extra information being used to try to reveal identity	**Determine, given the nature of the information to be published, and the threat level, whether the risk of extra information being used to try to reveal identity is "normal" or "high", and record your reasoning.** Many techniques may be used by someone intent on discovering new information about one particular person, or any person, within the data set to be released[21]. Numerous sources of information available now and in the future could conceivably be used, such as the electoral register, and social media sites. Thus it should be assumed that the means for re-

[21] For useful guidance on potential methods, see chapter 9 of "Security Engineering" by Ross Anderson, available at: http://www.cl.cam.ac.uk/~rja14/book.html

Step name	Process step description
	identification may be there, whenever data about people are being published, whatever the specific nature of that data. However, you may have special cause to give extra protection to the identity of individuals where there is: a) a skewed distribution of prevalence in population; b) special knowledge about individuals; or c) known availability of especially relevant information These three factors are most easily explained through examples. a) Skewed distribution: Sickle cell anaemia, for example, is not evenly distributed across the population. It is especially prevalent in people (or their descendants) from parts of tropical and sub-tropical regions where malaria is, or was, common. As a result, it is very roughly ten times more likely to occur in people from certain ethnic groups. Therefore, especially for populations where there are few people from the ethnic groups prone to the disease, the additional knowledge of the skewed distribution sickle cell anaemia can assist in the re-identification of individuals from published data about the disease. For example, publishing the fact that one person registered with Anytown General Practice has the disease could be very revealing in an area where residents almost exclusively self-classify their ethnic group as White British. b) Special knowledge: Consider a scenario where a national newspaper has published an expose of the private life of a famous actor, revealing where his children go to school, the care home of his elderly parent, where he shops, and many other personal details. With so much information in the public domain, information published about (say) the prevalence of disease in the actor's area of residence would be particularly vulnerable to re-identification. For example, it may mean that it is possible to associate the actor or members of his family with a particular disease (thereby revealing new confidential information). c) Known availability of especially relevant information: you can never be sure that no information exists that could be used in conjunction with the data set you propose to publish to re-identify individuals. However, you may discover that particularly relevant information does exist. Imagine you plan to publish data broken down by care home, and by age range. For two of the care homes, there is only one resident of age 60-65. Having searched the internet, you discover that the local hospital trust has published data on

Step name	Process step description
	discharges of patients with dementia, broken down into the care home into which they were discharged, and by age range. These data could be used together to discover that the two 60-65 year old patients in the two care homes have dementia – having re-identified the two residents, someone could learn something confidential about them (that they have dementia). As a result, you may decide that the existing publication by the trust is especially relevant information that introduces a relatively high risk of re-identification.[22] **If one of these three special factors (a, b or c) exists, assess the risk of extra information being used to try to reveal identity as "high".** **If the threat level is assessed (in the previous step) as "high", then assess the risk of extra information being used to try to reveal identity as "high".** **Otherwise (i.e. if neither is "high"), assess the risk of extra information being used to try to reveal identity as "normal".** The judgement may be difficult, and it is good practice to consult a colleague (or Caldicott Guardian – see below). Note that the above advice provides a "rule of thumb" re-identification risk level. It is not possible to predict risk precisely. Where it is particularly important to assess risk as accurately as possible (e.g. because the data being released is particularly sensitive, or there is a particularly great benefit to be derived from releasing as much information as possible) then specialist computer software may be purchased to assist with the assessment of re-identification risk.
Select anonymisation plan	**Based on the assessed risk of extra information being used to try to reveal identity, choose an anonymisation plan to reduce the intrinsic identifiability of the data to be released to the desired level.** Two tables are shown below. The first summarises a set of possible anonymisation plans where the risk of extra information being used to try to reveal identity is normal. These plans allow relatively more information to be released. The second table shows a summary of plans aimed at circumstances where the risk of extra information being used to try to reveal identity is high (thus allowing relatively little intrinsically identifiable information to be released).

[22] For more information on this subject, see the Anonymisation Code of Practice from the Information Commissioner's Office.

<table>
<tr><th>Step name</th><th>Process step description</th></tr>
<tr><td></td><td>
The tables below say “cells to be published relate to population”. The concept of “population” here is taken from GSS/GSR Disclosure Control Policy for Tables Produced from Administrative Data Sources[23], which is part of the National Statistician’s Guidance. It is best explained through example. For example, when reporting total abortions for Anytown, which has a total population of 100,000, the relevant “population” for the cell might be estimated at approximately 25,000 because males, young children and elderly women can be excluded (we know the cell only relates to the population of females of child-bearing age). A breakdown of abortion by age group for Anytown would mean that the relevant population for each cell was much smaller, e.g. there might be only 1,500 females of aged 10-15. However, suppose figures were broken down by ethnic group, and indicated that 5 women in Anytown who had had an abortion had self-classified as Irish. This is a special case, because the fact that a woman has had an abortion and her ethnic category (of Irish) both count as “sensitive personal data”[24], and so the relevant population should be the actual figure to be published in the cell – in this case “5” (and thus statistical disclosure control is almost certainly necessary).[25]

Although “population” can be difficult to assess (e.g. for a hospital), try to estimate ‘how many possible people could this be?”.

Notes that K-anonymity without automated software is only feasible with small numbers of records, and even then it is prone to error. If K-anonymity software is not available, it is often preferable to publish data in aggregated format.

Anonymisation plans to use where risk is normal
<table>
<tr><th>Plan no</th><th>Plan description</th></tr>
<tr><td>1</td><td>Where cells to be published relate to population > 1,000 people, derive aggregate data without statistical disclosure control.

Risk is normal, aggregated data</td></tr>
<tr><td>2</td><td>Where cells to be published relate to population ≤1,000 people, derive aggregate data with</td></tr>
</table>
</td></tr>
</table>

[23] Available at: http://www.ons.gov.uk/ons/guide-method/best-practice/disclosure-control-policy-for-tables/index.html
[24] What constitutes “sensitive personal data” is defined in the Data Protection Act. See: http://www.ico.gov.uk/for_organisations/data_protection/the_guide/key_definitions.aspx
[25] See pages 7-9 of “The exemption for personal information” at: http://www.ico.gov.uk/Global/faqs/~/media/documents/library/Freedom_of_Information/Detailed_specialist_guides/PERSONAL_INFORMATION.ashx

Step name	Process step description

	statistical disclosure control Risk is normal, aggregated data
3	Derive individual-level data to “weak” k-anonymity[26] Risk is normal, individual-level data

Anonymisation plans to use where risk is high

Plan no	Plan description
4	Where cells to be published relate to population > 10,000 people, derive aggregate data without statistical disclosure control Risk is high, aggregated data
5	Where cells to be published relate to population ≤10,000 people, derive aggregate data with statistical disclosure control Risks is high, aggregated data
6	Derive individual-level data to “strong” k-anonymity Risk is high, individual-level data

By their nature, individual-level data are much more likely to reveal an individual's identity than aggregate data. The risk of identifying an individual increases as the number of data items increases, and the risk that some individual becomes identifiable increases with the number of individual-level records. Virtually any data item in an individual-level data set may be a source of "outliers" - values that are unusual or unique that can reveal the identity of an individual either when used on their own or in conjunction with other data items. In order to address these risks, when releasing data into the public domain, very strong controls (such as provided through k-anonymity – see section 6.4.1) are necessary to prevent unique combinations of data items revealing a person’s identity and ensure that the anonymisation processing results in non-identifying data. To achieve the required k-anonymity level when releasing individual-level information into the public domain, it may be necessary to reduce the number of data items released and/or provide just a sample of records. Sampling has an added benefit in that an individual seeking to discover a person’s identity from the data cannot be sure

[26] For a definition of k-anonymity see the glossary. For more detail on the technique, see section 6.3.

Step name	Process step description
	whether that person is within the data set.
Refine anonymisation plan and specify anonymisation	**Having chosen an anonymisation plan, refine the plan to fit the circumstances of the release, and specify the anonymisation to be carried out. Confirm the plan with colleagues where appropriate, and record the plan and your reasoning.** The generic anonymisation sub-process steps described here for choosing a broad plan have to be practical and relatively easy to follow. The standard assessments on which the anonymisation plan is built are relatively simplistic e.g. assessments of "high" or "normal". Therefore, the anonymisation plan chosen should be reviewed to ensure it can be expected to provide a non-identifying data set. Furthermore, the standard anonymisation plans can be adjusted (e.g. to set k-anonymity to 8, or to adjust the degree of statistical disclosure control). Having reviewed the standard plan and considered the circumstances and expected data to be output, the plan should be refined if necessary, and the anonymisation processing specified. In practice, trial and error may work best: apply the initial anonymisation plan to the data set and review the output to refine the plan. Sometimes judgements may have to be made about data items that identify a person but do not reveal confidential information. Data that reveal confidential information about an identifiable person are not publishable. Data that identify a person, but do not reveal anything confidential, are often publishable. For example, publishing identifiers of a health care professional or other members of staff does not necessarily reveal anything confidential, and so it may not be necessary to anonymise the data prior to publishing (e.g with k-anonymity), although in such cases, the decision (ultimately for the SIRO or Caldicott Guardian) must be justifiable in the public interest. If you are considering publishing identifying information about staff, see section 2.7 of the Implementation Guidance. Aggregate tables about a general practice may be publishable without obscuring small numbers despite the small population base if they reveal that one person in the practice has dementia, but no other information about the patient (as in the case of the Quality Outcomes Framework tables published in recent years). If in doubt about how much to publish, bear in mind that the impact of releasing too little information tends to be less damaging than releasing too much. In border-line cases where the consequences of re-identification would be significant –

Step name	Process step description
	because, for example, they might leave an individual open to damage or distress, it may be expedient to err on the side of caution and restrict the information published. The Information Commissioner has said that he will take potential impact into account when considering complaints made by people who have been re-identified from a published data set. More information can always be released later, whereas information released cannot be withdrawn. However, if too much information is omitted, the publication may no longer be useful.
Consult Caldicott Guardian/SIRO	**Consult the Caldicott Guardian, Senior Information Risk Owner (SIRO) [27] and/ or confidentiality lead for the organisation on the action to be taken, and the required anonymisation specification (if any), and record the outcomes.** In a health and social care organisation without a Caldicott Guardian, another person with responsibility for information governance should be consulted. The Caldicott Guardian (or other responsible person) should consider the risk assessment and anonymisation specification, and be satisfied that data to be released are non-identifying data, and therefore that the risk of a person's identity being revealed from the data is insignificant. If the risk of re-identification is significant, then the anonymisation specification will require change. Consulting the Caldicott Guardian/SIRO is described as a relatively late step within this anonymisation process. In some circumstances, particularly where difficult judgements are needed at an earlier stage, staff may choose to also consult earlier in the process. For example, when planning to publish individual-level data, early discussion of potential sensitivities and risks may be helpful.

[27] Where the data relates to information other than patients, then the Senior Information Risk Owner should be consulted. See http://www.connectingforhealth.nhs.uk/systemsandservices/infogov/igap/igaf/roles.pdf

5 Implementation guidance: standard anonymisation plans

5.1 Introduction

A summary of the six standard anonymisation plans identified in section 4 are specified below, with one table for each of the standard anonymisation plans. These plans are drawn directly from the plans set out briefly in the sub-process description of "Select anonymisation plan" in section 4.3. The first three plans apply where the re-identification risk has been assessed as normal, and the final three plans apply when the risk has been assessed as high.

Each table below lists the mandatory techniques that form part of the standard anonymisation plan. This specification of techniques may be qualified, or explained further, in a free-text comment at the bottom of each anonymisation plan table.

No description of the techniques in each plan is given in the anonymisation plan tables below; for the specification of the techniques identified (such as "k-anonymity", and "statistical disclosure control"), see section 6.

Some of the tables refer to "underlying population" – see the step "Select Anonymisation Plan" in section 4.3 for an explanation of this concept.

5.2 Standard anonymisation plans when re-identification risk is normal

5.2.1 Plan 1: Where cells to be published relate to underlying population > 1,000 people, derive aggregate data without statistical disclosure control

Mandatory techniques	Aggregation
Qualifying comment	This is identical to plan 4 below, except that this requires that no cell released relates to a population size of under 1,000 people. As a consequence, it results in information that is more intrinsically identifiable than with plan 4 (all other things being equal).

5.2.2 Plan 2: Where cells to be published relate to underlying population ≤1,000 people, derive aggregate data with statistical disclosure control

Mandatory techniques	Aggregation Statistical disclosure control
Qualifying comment	This is identical to plan 5 below, except that this requires that no cell released relates to a

	population size of under 1,000 people. As a consequence, it results in information that is more intrinsically identifiable than with plan 5 (all other things being equal).

5.2.3 Plan 3: Derive individual-level data to "weak" k-anonymity

Mandatory techniques	Suppression of direct identifiers Reduction in detail in indirect identifiers (such as date of birth, postcode) "Weak" k-anonymity
Qualifying comment	This is identical to plan 6 below, except that k-anonymity control is weaker. As a consequence, it results in information that is more intrinsically identifying than with plan 6 (all other things being equal). "Weak" k-anonymity is where: - K = 3 - the variables not controlled through k-anonymity must exclude: o any derivation of date of birth (such as age range) o gender o ethnic category o any derivation of postcode (such as area code) o event dates (such as hospital admission date, whereas hospital admission month and year is acceptable) o employer o occupation or staff group Note that in order to achieve k-anonymity, additional anonymisation techniques may be necessary[28].

[28] Which techniques will depend on the circumstances of the case, and the judgement of those involved, as to which is likely to achieve the target level of k-anonymity with the smallest loss of value in data content. For this reason, a standard set of additional techniques on top of those listed is not specified.

5.3 Standard anonymisation when re-identification risk is high

5.3.1 Plan 4: Where cells to be published relate to underlying population > 10,000 people, derive aggregate data without statistical disclosure control

Mandatory techniques	Aggregation
Qualifying comment	This is identical to plan 1 above, except that this requires that no cell released relates to a population size of under 10,000 people. As a consequence, it results in information that is less intrinsically identifiable than with plan 1 (all other things being equal) in order to address the increased re-identification risk.

5.3.2 Plan 5: Where cells to be published relate to underlying population ≤10,000 people, derive aggregate data with statistical disclosure control

Mandatory techniques	Aggregation Statistical disclosure control
Qualifying comment	This is identical to plan 2 above, except that this requires that no cell released relates to a population size of under 10,000 people. As a consequence, it results in information that is less intrinsically identifiable than with plan 2 (all other things being equal) in order to address the increased re-identification risk.

5.3.3 Plan 6: Derive individual-level data to “strong” k-anonymity

Mandatory techniques	Suppression of direct identifiers Reduction in detail in indirect identifiers (such as date of birth, postcode) “Strong” k-anonymity
Qualifying comment	This is identical to plan 3 above, except that k-anonymity control is stronger. As a consequence, it results in information that is less intrinsically identifying than with plan 3 (all other things being equal), in order to address the higher re-identification risk.

	"Strong" k-anonymity is where: - K = 5 - all variables but one must be controlled through k-anonymity - the uncontrolled variable should not be full postcode, date of birth, or ethnic category[29]. Note that in order to achieve k-anonymity, additional anonymisation techniques may be necessary[30].

[29] The variables listed are excluded, and care should be taken before any other indirect identifier is chosen as the uncontrolled variable because indirect identifiers are more likely to risk re-identification.
[30] Which techniques will depend on the circumstances of the case, and the judgement of those involved, as to which is likely to achieve the target level of k-anonymity with the smallest loss of value in data content. For this reason, a standard set of additional techniques on top of those listed is not specified.

6 Implementation guidance: standard anonymisation techniques

6.1 Introduction

Each of the six standard anonymisation plans specified in section 5 comprises a set of standard anonymisation techniques. Those techniques are specified here in section 6. Anonymisation plans 1, 2, 4 and 5 involve deriving aggregate data (see section 6.2), and plans 3 and 6 involve deriving individual-level data (see section 6.3).

Each standard anonymisation technique in section 5 is described relatively briefly in a table setting out what transformation the technique is intended to achieve. This is followed by a more detailed description. Some of these standard techniques are complex and/or detailed, and most have been the subject of publications elsewhere. Wherever possible, existing authoritative published descriptions and specifications are selected and used here to provide the standard technique (to avoid needless re-invention). Each of the standard anonymisation techniques identified and described in this section appears as an element of one or more of the standard anonymisation plans set out in section 5. It is not intended to be a comprehensive set of possible anonymisation techniques; details of many alternative anonymisation techniques have been published (but these are not part of this standard).

For each of the anonymisation techniques identified here as part of the standard, a table is provided which:

- Describes briefly the transformation achieved by the anonymisation technique;
- Any essential requirements of the technique if it is to meet the standard;
- Further information describing the technique, and/or a reference to further information about the technique.

6.2 Anonymisation techniques when deriving aggregate data

6.2.1 Aggregation

Transformation required	**Input**	**Processing**	**Output**
	Individual-level data	Summarising groups of records into table(s)	Aggregate data
Essential standard requirements	No cell in the table(s) output may relate to a single individual.		
Further information	None		

Further information (references)	None

6.2.2 Statistical disclosure control

The single table below describes a number of related techniques, any of which may be used to transform small numbers (e.g. less than '5') appearing in cells within tables of aggregate data.

<table>
<tr><td>Transformation required</td><td>
<table>
<tr><th>Input</th><th>Processing</th><th>Output</th></tr>
<tr><td>Aggregate data</td><td>A process to change or remove small cell values (one of several possibilities – see below)</td><td>Aggregate data without small cell values</td></tr>
</table>
</td></tr>
<tr><td>Essential standard requirements</td><td>No cell in the aggregate data output may be ≤ 5.</td></tr>
<tr><td>Further information</td><td>Table Redesign
<table>
<tr><th>Description</th><th>Advantages</th><th>Disadvant-ages</th></tr>
<tr><td>Disguise unsafe cells, for example, by:
- grouping categories within a table
- aggregating to a higher level geography or for a larger population sub-group
- aggregating tables across a number of years/months/quarters
- rounding</td><td>Original counts in the data are not damaged
Easy to implement</td><td>Detail in the table will be reduced
May be policy or practical reasons for requiring a particular table design</td></tr>
</table>
If unsafe cells remain in the output tabulation, further protection methods should be considered in order to disguise them. The recommended method for post-tabular protection for most frequency tables is controlled rounding. In some cases, if the number of unsafe cells is low then cell suppression can be an alternative method. Controlled rounding and cell suppression can be</td></tr>
</table>

implemented in the Tau-Argus software[31].

Cell Modification – cell suppression

Description	Advantages	Disadvantages
Unsafe cells are not published. They are suppressed and replaced by a special character, such as '..' or 'X', to indicate a suppressed value. Such suppressions are called primary suppressions. To make sure that the primary suppressions cannot be derived by subtraction, it may be necessary to select additional cells for secondary suppression	Original counts in the data that are not suppressed are not adjusted Can provide protection for zeros	Most of the information about suppressed cells will be lost. Secondary suppressions will hide information in safe cells Information loss will be high if more than a few suppressions are required In order to protect any disclosive zeros, these will need to be suppressed. Does not protect against disclosure by differencing Complex to implement optimally if more than a few suppressions are required, and particularly complex for linked tables.

Cell Modification – rounding

[31] Available at http://neon.vb.cbs.nl/casc

	<table><tr><th>Description</th><th>Advantages</th><th>Disadvant-ages</th></tr><tr><td>Rounding involves adjusting the values in all cells in a table to a specified base[32]. This creates uncertainty about the real value for any cell while adding a small amount of distortion to the data.</td><td>Counts are provided for all cells Provides protection for zeros Protects against disclosure by differencing and across linked tables Controlled rounding preserves the additivity of the table and can be applied to hierarchical data</td><td>Cannot be used to protect cells that are determined unsafe by a rule based on the number of statistical units contributing to a cell Random rounding requires auditing; controlled rounding requires specialist software.</td></tr></table> **Cell modification - Barnardisation** <table><tr><th>Description</th><th>Advantages</th><th>Disadvantage s</th></tr><tr><td>A post-tabular method for frequency tables where internal cells of every table are adjusted by +1, 0 or -1, according to probabilities</td><td>Protects against disclosure by differencing</td><td>High level of adjustment may be required in order to disguise all unsafe cells Will distort distributions in the data</td></tr></table>
Further information (references)	The “further information” directly above was extracted directly from section 5 (“Selecting disclosure control methods”) of the *GSS/GSR Disclosure Control Policy for Tables Produced from Administrative Data Sources*[33],

[32] For more information about rounding and bases, see: http://en.wikipedia.org/wiki/Rounding
[33] Available at: http://www.ons.gov.uk/ons/guide-method/best-practice/disclosure-control-policy-for-tables/index.html

	which is part of the National Statistician's guidance. More information is available in that document.

6.3 Anonymisation techniques when deriving individual-level data

6.4 Data Suppression

Transformation required	**Input**	**Processing**	**Output**
	Individual-level data	Identifying a sample of records, or certain data items in all records, and withholding these from the output.	Individual-level (but fewer) data
Essential standard requirements	Free-text data items, and human images, must be suppressed. No other fixed standard test, but data reduction must be sufficient to make a significant contribution to anonymisation.		
Further information	Variable Suppression - This technique involves the removal or withholding of a data variable's values (e.g. removing name, address, postcode from an output) - All other variables in the record, i.e., those that are not quasi-identifiers, remain untouched - It may not always be plausible to suppress some variables because that will reduce the utility of the data Record Suppression - If variable suppression and reduction in detail techniques do not adequately anonymise the data then the alternative is the removal and withholding of the data records that create a high re-ID risk - However extensive suppression can introduce a high level of distortion in some types of analysis since the loss of records is not completely random and may reduce the usefulness		
Further information	The "further information" directly above was extracted directly from section 12.2 of "'Best Practice' Guidelines		

(references)	for Managing the Disclosure of De-Identified Health Information"[34], published by the Canadian Institute for Health Information. Further information is also given in section 3.1 of "Tools for De-Identification of Personal Health Information"[35], published by the Pan Canadian Health Information Privacy (HIP) Group.

6.4.1 k-anonymity ("strong" and "weak")

Transformation required	**Input**	**Processing**	**Output**
	Individual-level data	Identify matching sets of records (see further information below)	Individual-level data
Essential standard requirements	Weak anonymity: - k = 3 - the variables (i.e. data items) not controlled through k-anonymity must exclude: o any derivation of date of birth (such as age range) o gender o ethnic category o any derivation of postcode (such as area code) o event dates (such as hospital admission date, whereas hospital admission month and year is acceptable). o employer o occupation or staff group Strong anonymity: - k=5 - all variables but one must be controlled through k-		

[34] See: www.ehealthinformation.ca/documents/de-idguidelines.pdf
[35] See: https://www2.infoway-inforoute.ca/Documents/Tools_for_De-identification_EN_FINAL.pdf

<table>
<tr><td></td><td>anonymity
- the uncontrolled variable should not be full postcode, date of birth, or ethnic category[36].</td></tr>
<tr><td>Further information</td><td>A data set provides k-anonymity for the data subjects represented if the information for each person contained in the data set cannot be distinguished from at least k-1 individuals whose information also appears in the data set. For example, a data set has 5-anonymity if, for every record in the data set that describe characteristics of a data subject, there are at least four other individuals also represented by records in the data set who share the same characteristics described by the record.

The following record-level data set exhibits 3-anonymity:

Example of K-Anonymity where K=3

<table>
<tr><th>Record no</th><th>Age range</th><th>Gender</th><th>ICD-10 code</th></tr>
<tr><td>1</td><td>0 to 10</td><td>M</td><td>F106</td></tr>
<tr><td>2</td><td>20 to 35</td><td>F</td><td>F106</td></tr>
<tr><td>3</td><td>0 to 10</td><td>M</td><td>F106</td></tr>
<tr><td>4</td><td>51 to 65</td><td>F</td><td>F106</td></tr>
<tr><td>5</td><td>20 to 35</td><td>M</td><td>F106</td></tr>
<tr><td>6</td><td>51 to 65</td><td>F</td><td>F106</td></tr>
<tr><td>7</td><td>0 to 10</td><td>M</td><td>F106</td></tr>
<tr><td>8</td><td>20 to 35</td><td>F</td><td>F106</td></tr>
<tr><td>9</td><td>51 to 65</td><td>F</td><td>F106</td></tr>
<tr><td>10</td><td>20 to 35</td><td>F</td><td>F106</td></tr>
<tr><td>11</td><td>20 to 35</td><td>M</td><td>F106</td></tr>
<tr><td>12</td><td>20 to 35</td><td>M</td><td>F106</td></tr>
<tr><td>13</td><td>0 to 10</td><td>M</td><td>F106</td></tr>
</table>

Typically, k-anonymity control is designed to operate on only a subset of variables in the data set for release (such as those related to postcode, gender and dates). Other variables, such as diagnosis and hospital admission date typically are not controlled. This is the approach taken</td></tr>
</table>

[36] The variables listed are excluded, and care should be taken before any other indirect identifier is chosen as the uncontrolled variable because indirect identifiers are more likely to risk re-identification.

	here for "weak k-anonymity". However, that leaves open the possibility that the set of data items outside the controlled group can reveal identity. For example, if diagnosis and hospital admission date are outside the control group, then a person who knows that their neighbour was admitted on a certain date, may be able to deduce from the district postcode, age range and gender that the only person that could be is his neighbour. He would then discover his neighbour's diagnosis. The risk posed from this is likely to be negligible when the value and interest is low, but when interest is high (e.g. when publishing HIV data), then "strong" k-anonymity is appropriate. With "strong" k-anonymity, a maximum of one data item is allowed outside the control group, and so the probability of a person's identity being revealed is nearly zero, and the chance of learning some new confidential information about a person is small. One reason why k-anonymity is chosen here as a standard is that it is implementable, and forms the underlying basis of a number of existing software products for protecting privacy.
Further information (references)	Some of the above explanation was extracted from section 2.5 of "Tools for De-Identification of Personal Health Information"[37], published by the Pan Canadian Health Information Privacy (HIP) Group. Further information on k-anonymity is also available in section 5 of that document, and in section 16 of "'Best Practice' Guidelines for Managing the Disclosure of De-Identified Health Information"[38], published by the Canadian Institute for Health Information, and in the HIPAA de-identification guidance from the USA[39]. Many papers promoting and critically reviewing k-anonymity have been published in peer-reviewed academic journals[40].

[37] See: https://www2.infoway-inforoute.ca/Documents/Tools_for_De-identification_EN_FINAL.pdf
[38] See: www.ehealthinformation.ca/documents/de-idguidelines.pdf
[39] See: http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/guidance.html
[40] One article advocating k-anonymity control can be found at: www.epic.org/privacy/reidentification/Sweeney_Article.pdf An academic article identifying weaknesses of k-anonymity (and especially "weak" k-anonymity) can be found at: www.cs.colostate.edu/~cs656/reading/ldiversity.pdf

6.4.2 Reduction in detail in indirect identifiers (such as date of birth, postcode)

Transformation required	**Input**	**Processing**	**Output**
	Individual-level data	Identify, and withhold or transform indirect identifiers so they are less likely to reveal identity	Individual-level data without indirect identifiers, or with indirect identifiers
Essential standard requirements	Post code truncated to either area code or district code[41]. No date of birth (e.g. transform to age, year of birth, or 5-year age band). No event dates (e.g. transform admission date to admission year, or month and year).		
Further information	None		
Further information (references)	None		

6.4.3 Suppression of direct identifiers

Transformation required	**Input**	**Processing**	**Output**
	Individual-level data	Identify and withhold direct identifiers	Individual-level data without direct identifiers
Essential standard requirements	Suppression of name, address, widely-used unique person or record identifier (notably National Insurance Number, NHS Number, Hospital Number), telephone number, email address, and any other data item that on its own could uniquely identify the individual.		
Further information	None		
Further information (references)	None		

[41] For an explanation of area and district postcode, see: http://www.geoplan.com/GIS_Mapping/How_Postcodes_Work. Note that the distribution of postcode elements is not uniform, and that there are very few individuals living in some districts. Therefore, particular care should be taken when publishing data about sparsely-populated geographical areas.