**Corporate policy**

**Anonymisation and Pseudonymisation Standard**

**Issue sheet**

|  |  |
| --- | --- |
| Document reference | NHSBSADPN005a |
| Document location | S:\BSA\IGM\Mng IG\Developing Policy and Strategy\Develop or Review DPA Policy\Current and Final |
| Title | NHS Business Services Authority Anonymisation and Pseudonymisation Standard |
| Author | Chris Gooday |
| Issued to | All NHSBSA staff |
| Reason issued | For information / action |
| Last reviewed | 15 April 2019 |
| Date of Equality Assessment | N/A |
| Date of Fraud Review | N/A |

**Revision details**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Version** | **Date** | **Amended by** | | **Approved by** | **Details of amendments** |
| Initial release | 15.11.2017 | CG | APF | | Restructure wording from previous Data Protection and Freedom of Information Polices to an ISMS Standard |
|  | 25.07.2018 | CG | GDPR | | Amended as name, address, postcode and DOB can normally only be used as an identifier when combined together |
|  | 15.04.2019 | CG |  | | Reflect audit findings |
|  | 07.01.2020 | CG |  | | Reflect review of Statistical Disclosure Controls |

**Scope**

This standard covers how the risks of the identification of individuals and organisations from information are considered

It can be applied to quantitative data as well as qualitative data such as free text comments.

1. **Anonymisation and Pseudonymisation Standard**
   1. This standard supports and acts as a measure of compliance with both the [Freedom of Information Policy](https://www.nhsbsa.nhs.uk/sites/default/files/2019-02/NHSBSAFOI001d%20NHSBSA%20Freedom%20of%20Information%20Policy.pdf) and the [Data Protection and Confidentiality Policy](https://www.nhsbsa.nhs.uk/sites/default/files/2019-02/NHSBSA%20Data%20Protection%20and%20Confidentiality%20Policy.pdf).
   2. This standard will comply with the legal and professional obligations set out in the:

## [ICO Anonymisation Code of Practice](https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf)

* 1. Be consistent with and support the Statistical Disclosure Controls Protocol [link]
  2. Use the minimum data needed to achieve the business objective.
  3. Apply Anonymisation or Pseudonymisation techniques as early as possible.
  4. When Personal Data is being used for purposes not related to the reason we received it, you need to justify to your peers why and when the individual needs to be identifiable and document this. For example, there may be no common key to match different data sets and fuzzy logic is required so identifiable personal data is justified.

1. **Anonymisation**

Anonymisation is the process of removing identifiers, both direct and indirect, that may lead to an individual or organisation being identifiable. Once data is truly anonymised and individuals are no longer identifiable

2.1 Where tanonymisation data use is appropriatley the following Anonymisation standard will be applied:

* Remove all direct identifiers such as NHS Number, National Insurance Number and organisation name.
* Remove all well-known indirect identifiers such as name, address, postcode, date of birth.
* Assess the risk of identification from any remaining indirect identifiers as detailed in the Statistical Disclosure Control Protocol.
* Consider this as a disclosure to anyone in the world, and assess the risks of re-identification by a) a motivated individual b) the individual/organisation c) the NHSBSA .
* Assess the harm that could be caused to the individuals/organisations if the data were to be re-identified?
* Assess the current information that is publicly available and also likely to be available to the receiptinet of the information that could be linked to this information to identify the individual/organisation.
* Consider how technology has changed to use increased computing power and sophisticated techniques re-identify the information?
* Be aware that unauthorised re-identification is a criminal offense.
* Reduce the risk of re-identification to that which is low/remote by applying the Statistical Data Control protocol:

2.2 Keep a summary of the techniques used prior to any disclosure.

2.3 Consider the impact of applying the standard on the usability of the anonymised data.

1. **Pseudonymisation**

Pseudonymisation disguises or masks the identity of the individual or organisation the data relates to. However, they can still be identified by reversing the Pseudonymisation process.

3.1 When the individual only needs to be identified in a limited set of circumstances Pseudonymisation of the data can be used. . This means that the identity of the individual will be disguised for the majority of processing. Only where the data shows unusual patterns that justify identification will the individual be identified. For example, to investigate fraud or a patient safety issue. The following Pseudonymisation standard will be applied:

* Pseudonymisation techniques should be used wherever possible to comply with the GDPR principle of privacy by default and design. This ensures access to identifiable personal data is on a need to know basis.
* Pseudonymised data is still personal data that has been made more secure through the use of pseudonymisation. Encryption can be considered to be a form of pseudonymisation. Consequently:
  + Only display the pseudonymised data items that are required.
  + Apply the same security measures as for personal data.
  + Ensure that a data sharing agreement and a Data Protection Impact Assessment have been completed prior to sharing with a third party.

3.2 Carefully consider the feasibility of Pseudonymisation and consider factors such as:

* The impact on performance of the pseudonymisation process as this carries a processing overhead.
* Confirm with the software supplier if the information technology employed supports pseudonymised data.
* The quality of the data is sufficient to ensure that reversing of the Pseudonymisation is not regularly required to check the integrity of the pseudonymised data?

3.3 The Pseudonymisation algorithm must be:

* Consistent across the data sets it will be used infor an agreed period of time.
* Checked to assure the integrity of the pseudonymised data.
* Strong enough to prevent reverse engineering during the lifetime of the data.

Any change in the algorithm needs to be risk assessed and justified.

3.4 The algorithm used must have strong management controls in place including:

* The Pseudonymisation keys must be stored in a database encrypted to [NHS Encryption Standards](https://digital.nhs.uk/binaries/content/assets/legacy/pdf/n/j/encryption_-_good_practice_guide_230517.pdf).
* Access must be restricted to a minimal number of authorised staff.
* There must be clear agreed rules for each data set as to when reversal of the pseudonymisation process is justified.
* Having a clear process as to how the algorithm would be changed if the keys were compromised.
* Where a key is shared outside the NHSBSA then there must be a clear agreement about responsibilities to ensure this standard is adhered to.
* A destructive Pseudonymisation technique should be considered when disclosing to third parties. This only pseudonymises the data where you cannot replicate the resulting Pseudonymisation key.

3.5 Ideally, pseudonymisation should take place in the data store and consideration should be given when creating a data warehouse to applying a standard pseudonymisation technique which would prevent all users having access to direct or aggregated identifiers where possible. If this is not possible then the information reported should be pseudonymised where possible.

This application would require distinct data sets, one pseudonymised, one non-pseudonymised – to allow for identification where justified.

This would prevent unauthorised access to sensitive data internally and reduce the risk of data sharing breaches.

1. **Review** 
   1. This standard will be reviewed annually or on becoming aware that the re-identification risks or perceived impact on data subjects have changed.