Operational

Data Anonymisation and Pseudonymisation

Guidance for staff and students

This guidance covers the anonymisation and pseudonymisation of personal data for business or research purposes. If you are collect, store, process or transfer personal data for either business or research you need to be aware of the approaches outlined here, why they are required, where they applicable and who to contact if you have any concerns or questions relating to them.

The issue

Data protection law regulates how universities and other learning providers collect and use information about students, staff and others. It also provides individuals with the right to access information that is held about them. Personal data that directly, or indirectly in combination with other data sources identifies a living individual and is governed in the UK by the Data Protection Act (2018) and General Data Protection Regulation (GDPR) [PDF]. These laws require increased accountability and transparency from all those that collect and handle any information relating to an identifiable individual (personal data).

Data protection failures are now regularly headline news and so any significant failings by staff or students within the School could not only trigger data protection liability with financial repercussions but impact the reputation of the School as well. The risks of enforcement action and data subject claims can be mitigated through a range of practical and technical measures that include anonymisation and pseudonymisation technologies.

Key principles

Article 5 of the GDPR sets out key principles which lie at the heart of the general data protection regime. In brief personal data must be:

- Processed lawfully, fairly and transparently
- Collected only for specified purposes
- Limited to what is necessary for those purposes
- Kept accurate
- Held for no longer than is necessary
- Retained securely

What is Personal Data?

Personal data is defined in the GDPR as:

> "'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person".

This means personal data has to be information that relates to an individual. That individual must be identified or identifiable either directly or indirectly from one or more identifiers or from factors specific to the individual.

The GDPR covers the processing of personal data in two ways:

- personal data processed wholly or partly by automated means (that is, information in electronic form); and
- personal data processed in a non-automated manner which forms part of, or is intended to form part of, a 'filing system' (that is, manual information in a filing system).

In most circumstances, it will be relatively straightforward to determine whether the information you process 'relates to' an 'identified' or an 'identifiable' individual. In others, it may be less clear and you will need to carefully consider the information you hold to determine whether it is personal data and whether the GDPR applies.

This guidance will explain the factors that you should consider to determine whether you are processing personal data. These are:

- identifiability and related factors;
- whether someone is directly identifiable;
- whether someone is indirectly identifiable;
- the meaning of 'relates to'; and
- when different organisations are using the same data for different purposes.

The categories of personal data

Some of the personal data you may process can be more sensitive in nature and therefore requires a higher level of protection. The GDPR refers to the processing of these data as 'special categories of personal data'. This means personal data about an individual's:

- race;
- ethnic origin;
- political opinions;
- religious or philosophical beliefs;
- trade union membership;
- genetic data;
- biometric data (where this is used for identification purposes);
- health data;
- sex life; or sexual orientation.

Personal data can include information relating to criminal convictions and offences. This also requires a higher level of protection.

What about unstructured paper records?

Under the Data Protection Act 2018 (DPA 2018) unstructured manual information processed by a public authority such as the LSE constitutes personal data. This includes paper records that are not held as part of a filing system. There are several parts of the GDPR that do not apply to unstructured manual information, but it is better to anonymise or pseudonymise this data as well.

What is pseudonymised data and is it still personal data?

Pseudonymisation is a technique that replaces or removes information in a data set that identifies an individual.

The GDPR defines pseudonymisation as:

> "…the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person."

Pseudonymisation may involve replacing names or other identifiers which are easily attributed to individuals with, for example, a reference number. Whilst you can tie that reference number back to the individual if you have access to the relevant information, technical and organisational measures are put in place to ensure that this additional information is held separately.

Pseudonymising personal data can reduce the risks to the data subjects and help you meet your data protection obligations. However, pseudonymisation is effectively only a security measure. It does not change the status of the data as personal data. Recital 26 makes it clear that pseudonymised personal data remains personal data and within the scope of the GDPR.

> "…Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person…"

[See relevant provisions in the GDPR - Article 4(1), Article 4(5) and Recitals 26, 28 and 29].

What is anonymised data?

The GDPR does not apply to personal data that has been anonymised. Recital 26 explains that:

> "…The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes."

This means that personal data that has been anonymised is not subject to the GDPR. Anonymisation can therefore be a method of limiting your risk and a benefit to data subjects too. Anonymising data wherever possible is therefore encouraged. However, you should exercise caution when attempting to anonymise personal data.

It is worth noting that organisations frequently refer to personal data sets as having been 'anonymised' when, in fact, this is not the case. You should therefore ensure that any treatments or approaches you take truly anonymise personal data. The risk is that you may disregard the terms of the GDPR in the mistaken belief that you are not processing personal data. In order to be truly anonymised under the GDPR, you must strip personal data of sufficient elements that mean the individual can no longer be identified. However, if you could at any point use any reasonably available means to re-identify the individuals to which the data refers, that data will not have been effectively anonymised but will have merely been pseudonymised. This means that despite your attempt at anonymisation you will continue to be processing personal data. You should also note that when you do anonymise personal data, you are still processing the data at that point.

For a very thorough checklist on anonymising and/or pseudonymising data, the US Bureau of the Census provides the following checklist: https://www.census.gov/srd/sdc/drbchecklist51313.docx.

What about information about companies?

Information concerning a 'legal' rather than a 'natural' person is not personal data. Consequently, information about a limited company or another legal entity, which might have a legal personality separate to its owners or directors, does not constitute personal data and does not fall within the scope of the GDPR. Similarly, information about a public authority is not personal data. However, the GDPR does apply to personal data relating to individuals acting as sole traders, employees, partners, and company directors wherever they are individually identifiable and the information relates to them as an individual rather than as the representative of a legal person. A name and a corporate email address clearly relates to a particular individual and is therefore personal data. However, the content of any email using those details will not automatically be personal data unless it includes information which reveals something about that individual or has an impact on them.

How do we ensure privacy?

Access to student data and analytics should be restricted to those identified as having a legitimate need to view them. Where data is to be used anonymously particular care must be taken to avoid:

- Identification of individuals from metadata
- Re-identification of individuals by aggregating multiple data sources

The use of "special category data" for the purposes of learning analytics requires additional safeguards. Circumstances where data and analytics could be shared externally (e.g. requests from educational authorities, security agencies or employers) must be made explicit to staff and students and may require additional consent. The use of "special category data" for the purposes of learning analytics requires additional safeguards. Departments should ensure that student data is protected when third parties are contracted to store data or carry out learning analytics on it.

Researchers know they need to anonymise sensitive data where relevant. There are processes that they abide by in terms of making ethical use of data, alongside compliance with Freedom of Information (FoI) legislation and both the EU General Data Protection Regulation and the new UK Data Protection Act (2108). The latter two require more explicit consent for reuse of lawfully held personal data.

Questions and Concerns

If you are have any questions or concerns relating the use of personal data for research purposes please contact the School's Data Protection Officer Rachael Maguire (glpd.info.rights@lse.ac.uk).

Further reading

CESSDA ERIC (2018) Processing personal data

Finnish Social Science Data Archive Anonymization and Identifiers

Irish Qualitative Data Archive (2008) Anonymisation Guidelines [PDF]

Information Commissioners Office (2012)Anonymisation: Managing Data Protection [PDF]

Risk Code of Practice JISC Legal (2014) Data Protection and Research Data: Questions and Answers

UK Anonymisation Network (2016) Anonymisation Decision-Making Framework

Managing and Sharing Research Data: A Guide to Good Practice (2019) Corti, L et. al. (October 2019) 368pp. SAGE Publications Ltd

Recent article from JISC: https://www.jisc.ac.uk/guides/cloud-computing-innovation-gdpr-and-you

ICO Anonymisation Code of Practice: https://ico.org.uk/media/1061/anonymisation-code.pdf

ICO Article on Privacy Techniques and AI: https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2019/08/data-minimisation-and-privacy-preserving-techniques-in-ai-systems/

Review schedule

Review interval	Next review due by	Next review start
3 years	1/1/2023	1/10/2022

Version history

Version	Date	Approved by	Notes
0.1	03/11/2019	Draft for review only	
1.0	01/06/2020	IGMB	

Links

Reference	Link

Contacts

Position	Name	Email	Notes
Data Protection Officer	Rachael Maguire	r.e.maguire@lse.ac.uk	

Communications and Training

Will this document be publicised through Internal Communications?	**Yes/ No**
Will training needs arise from this policy	**Yes/ No**
If Yes, please give details	