Section 3:
Data Protection Principles and Concepts

This section discusses the data protection principles and key new concepts of accountability, data protection by design and default and pseudonymisation. It explains how these principles should be embedded through the research cycle.

3.1 Data protection principles

The GDPR, sets out six data protection principles, which largely cover the eight data protection principles set out in the DPA 1998:

- **Lawfulness, fairness and transparency** – Personal data must be processed lawfully, fairly and in a transparent manner.
- **Purpose limitation** – Personal data must be obtained for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes. Further processing is allowed for archiving, scientific, statistical and historical research purposes.
- **Data minimisation** – Personal data processed must be adequate, relevant and limited to what is necessary.
- **Accuracy** – Personal data must be accurate and, where necessary, kept up to date.
- **Storage limitation –** Personal data must not be kept longer than is necessary (but data processed for archiving, scientific, statistical and historical research purposes can be kept longer subject to safeguards).
- **Integrity and confidentiality** – Appropriate technical and organisational measures must be put in place to guard against unauthorised or unlawful processing, loss, damage or destruction.

A table illustrating the differences, between the GDPR and the DPA 1998 as set out in the Explanatory Memorandum for the UK Data Protection Bill is set out in the Appendix to this Guidance.

3.2 New data protection concepts

The new concepts of accountability, data protection by design and default and pseudonymisation work with the data protection principles to underpin the new legislation.

Accountability

Accountability requires that data controllers and data processors are responsible for, and are able to demonstrate compliance with the data protection principles. It requires research organisations to put in place appropriate technical and organisational measures and to be able to demonstrate what they did and its effectiveness.

Accountability measures include:

- Use of data protection impact assessments for high risk processing
- Appropriate documentation including internal records of processing activities
- Mandatory data breach notification regime
- Appointment of data protection officer

Data mapping, the process of identifying, understanding and mapping out the data flows of an organisation is a valuable process to support privacy compliance and underpin accountability. Data flows may vary project by project. These issues are discussed and explored further in Section 9 (Data Governance and Accountability) of Part 2 of the MRS Guidance (forthcoming July 2018).

Data protection by design and default

Data protection by design and default means that all data collection exercises must be proactively designed and conceptualised in the most privacy enhancing way. This needs to be done by embedding privacy in organisational practices, policies and procedures and can include:

- Limiting access to personal data– Ensure that only those who need access to data are granted access privileges.
- Minimising data collection – Limit data collection to the data required for the research project/exercise.
- Retaining personal data for reasonable but generally short periods – establish appropriate retention period(s), advise clients as to what the retention period is and periodically review and revise limits.

Alongside these organisational processes, technical safeguards and design systems of any IT architecture need to embed privacy. This must include data protection impact assessments (DPIA's) for applicable projects. DPIAs have a vital role to play in any GDPR compliance programme as they allow identification of potential privacy issues at an earlier and less costly stage. They also reduce the risks and increase of awareness of privacy and data protection with staff members throughout organisations.

In implementing data protection by design the GDPR requires that data controller(s) take into account several factors:

- the state of the art (which varies over time and is based on constantly evolving best practices and technology)
- the cost of implementation
- the nature, scope, context and purposes of processing
- the likelihood and severity of risks to the rights and freedoms of natural persons posed by the processing of their personal data.

This means that organisations can take a flexible risk based approach, the higher the risk, the more rigorous the measures that must be taken and as with other compliance obligations under the GDPR it will be key to ensure this is documented.

Pseudonymisation

Pseudonymisation of personal data is highly encouraged in processing data for research purposes. It is the processing of personal data so that it can no longer be attributed to a specific data subject without the use of additional information, such as a unique identifier, which can make the data identifiable. Although pseudonymised data is still personal data, it is a useful data security measure that can limit an organisation's risk profile and exposure for personal data breaches.

In order to become pseudonymised data, the unique identifier must be kept separately and held subject to adequate technical and organisational measures. Data can be considered as pseudonymised even where the unique identifier is kept within the same organisation. If the holder of the pseudonymised data does not have the means to reverse or unlock the pseudonymisation, then the data that they hold will be anonymised rather than pseudonymised data. The difference between pseudonymised and anonymised data is that for anonymised data there exists no key to link the data to the individual.

Although pseudonymised data may sometimes also be referred to as de-identified data, this is not a term that is used in the GDPR.

3.3 Embedding data protection principles and concepts in the research cycle

The data protection principles are inter-related and researchers need to ensure that the principles are followed, as applicable, throughout the full research cycle. Principles will need to be applied by all parties within the research supply chain to ensure that sub-contractors acting as data processors adhere to the policies and processes set out by the data controller(s), who may be the client and/or the lead researcher. Written contracts must always be used to clearly set out the roles and responsibilities of all parties within the research supply chain including the commissioning client, research agency, fieldwork agency as well as any freelance interviewers or recruiters.

This section sets out general points for consideration in applying the data protection principles to discrete research projects. Application to specific types of research is discussed in Section 7 Issues Relating to Specific Research (forthcoming July 2018)

Figure 1: Research Cycle

Scoping or setting-up research project

Researchers designing or setting up a data collection exercise must ensure that consideration is given to designing the research in such a way that the amount of personal data collected is only that which is necessary to meet the research objectives.

- Scope of what constitutes personal data must be broadly construed in light of the definition of personal data in the GDPR. This relates to the possibility of identifying an individual rather than a pre-defined list of information attributes. Consideration must always be given to the means and likelihood of re-identifying individuals. Re-identification risks are dynamic and will increase in line with technological improvements and reduced costs. In line with this, it is best practice to protect even "anonymised" special category data and ensure that it is stored securely in light of the higher level of harm if the data is re-identified.
- Researchers should always categorise photographs, audio recordings, video recordings and still images as personal data. The ease of technology in linking these to an identifiable person means that there is a higher risk of re-identification for this type of media. Transcripts of recordings can be used in order to properly anonymise audio and video recordings ensuring that the transcripts are edited to remove any comments that may lead to identification of a data subject in the research study. Alternatively, pixelating or blurring images can also be used to pseudonymise images. Photographs will also be special category data where the photos are used for the purposes of uniquely identifying a natural person such as in an electronic passport.
- Personal details or characteristics inferred or derived about data subjects from the analysis of data provided, rather than data provided directly by them, must also be treated as personal data or special category data as applicable. Observed data (such as online cookies automatically recorded), derived

data (such as data produced from using other datasets) or inferred data (such as using algorithms to predict health outcomes based on combining information in different datasets) produced during analysis of data may trigger different privacy risks than traditionally provided data that will need to be considered in assessing privacy implications of the research project. If special category data is inferred as a result of profiling, the data controller needs to make sure that the processing is not incompatible with the original purpose; there is a lawful basis for the processing of the special category data and data subject has been informed about the processing.[5] Researchers must ensure there is a processing ground for processing any special category data such as explicit consent where undertaking this type of project.

- Specific policy documentation requirements in UK DPA 2018 that apply to the collection of special category data or criminal convictions data must always be met.
- Consider the earliest point in the process that personal identifiers can be removed from any data in order to create pseudonymised or anonymised data.
- Review of the use of personal data in a research exercise may require the data controller(s) to undertake Data Protection Impact Assessments (DPIA) (previously known as a Privacy Impact Assessment (PIA)). A DPIA is only required when processing is likely to result in a high risk to the rights and freedoms of individuals. In these circumstances, it will be necessary to assess the risks and potential harm to data subjects such as where a project involves large scale collection of special category personal data or matching of datasets collected by different data controllers in a way that would exceed the reasonable expectations of individuals. DPIAs that identify high risk data collection exercises with risks that cannot be reduced or adequately mitigated by data controller(s) will require prior consultation with the ICO in line with published ICO timeframes.
- Submission of a research proposal to a client should include, where feasible, a data management plan that sets out the key data protection and privacy issues and makes suggestions for addressing any privacy issues and/or the necessity of a DPIA.

The legal ground that will be used to collect personal data such as consent of the data subject or the legitimate interests of the data controller in conducting the research must be identified and reviewed at the outset of the project to ensure that the processing is fair, lawful and transparent and that data subject rights can be met.

Collecting data

Researchers collecting personal data for a research exercise must:

- ensure the purpose of the data collection is clearly specified in an information notice (also known as a privacy notice or privacy information notice) which provides full details of all privacy information to data subjects[6]
- minimise the collection of personal data by only collecting data that is necessary
- ensure that data subjects are clearly informed about expected uses of data and provided with an adequate privacy information notice
- securely store and manage all data and build in security measures such as encryption or hashing of data taking into account the sensitivity of the data being collected and any risks to research data subjects.

[5] See A29 WP Guidance on Automated Individual Decision-Making and Profiling (6 February 2018)
[6] See MRS GDPR In Brief – Informed Consent for further details on information that must be included in information notices.

Analysing data

Researchers must bear in mind that analysis of personal data is also subject to the data protection principles. This is of equal importance to the analysis of data collected for research and the secondary use of existing data for research purposes:

- Personal data must only be used in accordance with privacy information notices provided to research data subjects.
- Use of aggregated data sets in quantitative projects and anonymised data in qualitative projects is preferable.
- Anonymised or pseudonymised datasets should be used as far as possible. It is also important for researchers to check and understand what other data research clients might have that could lead to identification of research data subjects being re-created from anonymised survey data in order to satisfy themselves that the data will remain as anonymised data upon transfer to a new environment e.g. a client's.
- Access to personal data must be limited to those researchers directly involved in the research exercise.

Reporting and/or publication

Personal data must not be included in research reports unless consent has been obtained from data subjects:

- Identifiable verbatim quotes from research data subjects can only be used with express consent of the data subject. Verbatims can be used without specific consent if they are anonymised (with the removal of identifying contextual detail and personal information).
- Visual images, particularly video clips, can be used if images are pixelated and/or voices disguised. Other visual images must only be used where data subjects have been fully informed about their use and have consented. Intention to use or share the video clips and/or images on websites or social media platforms must be made clear to the data subject in seeking their consent.

Retention and disposal

In line with the integrity and security data protection principle it is vital to:

- keep personal data secure and dispose of it securely taking into account the risk-level of any data
- store personal data in line with data retention policies
- ensure that all parties within the research data supply chain with access to personal data e.g. data processors follow the same processes and policies

For further information see:

- MRS GDPR In Brief (No. 5) Informed Consent (Member Content)
- ICO Guide to the GDPR
  https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/