UK Data Service

ukdataservice.ac.uk

Ethical and legal guidelines in data sharing – Q&A May 2025

08 May 2025

Public

Version: 01.00

“Can the participant have copyright in research data (e.g. interview transcript or recording)?”

Yes, they can have copyright. For example, without a clear assignment of copyright they are the copyright owner of their own recorded words in the interview or recording. However, the researcher holds the copyright of the transcripts. Copyright ownership needs to be agreed before the onset of the data collection. A range of copyright scenarios
outlined on our website might be helpful.

“What's the definition of (and criteria for) "effectively" anonymised data?”

The Information Commission’s Office (ICO), have published guidance of effectively anonymised data, as

“If in the consent form it does not specify re-use or sharing of the data, is the data allowed to be shared in a repository (when it's anonymous)?”

It is an ethical obligation to inform the participants about the future uses of the data. If this has not been stated in the consent form, it is advisable to obtain retrospective consent. However, there are situations where retrospective consent might not be possible but data sharing can be allowed under stricter access arrangements and with the approval of the ethical board that has approved the original research.

“Aren't legal and ethical aspects related but separate? Not sure if "policies of relevant organisations" would be quasi-legal ....”

Yes, legal and ethical aspects are related but separate. The legal aspects are those grounded in laws, regulations, and binding rules that must be followed and can be enforced by courts or regulatory bodies. In the context of research data sharing, these would be laws like UK GDPR, the Data Protection Act, or other regulations, which clearly define the rights of individuals and responsibilities of organisations when handling personal data.

On the other hand, ethical aspects deal with the moral principles that guide decision-making, often going beyond what the law mandates. For example, researchers may choose not to share data if they believe it could cause harm to individuals, even if the law permits it. Ethical guidelines in research often focus on principles like confidentiality, informed consent, and respect for privacy, which may not always be legally enforceable but are crucial for maintaining public trust and the integrity of research.

As for "policies of relevant organisations", these are usually internal guidelines created by institutions (universities, research organizations, etc.) to ensure their research complies with both legal requirements and ethical standards. However, these policies are not quasi-legal unless they are directly tied to laws or regulations that require compliance. They are important for ensuring that research is carried out responsibly, but they don't have the same legal weight as laws unless they are framed as part of legal compliance (e.g., policies related to the UK GDPR compliance).

“You said that there are special categories of personal data include things like ethnicity, religion, health among other things. However, these things on their own would not identify someone. E.g., if the name/DOB/address was removed from the data set, would these special categories still be considered 'personal data?”

Yes, even if information that directly identifies a participant such as the name, date of birth, and address are removed, information that might indirectly identify a participant which includes special categories of personal data, such as ethnicity, religion, or health information, can still be considered personal data under the UK GDPR, depending on the context. This is because these indirect identifiers, including information about ethnicity, religion, or health are pieces of information that when used together could lead to the identification of an individual. For example, if someone’s religion and health data are shared along with their job title and a low level geographic location such as a partial postcode or even a local authority, there might still be a risk of re-identifying them.

“Just a question for future reference: if you work in a team and there is disagreement about what data to share (e.g. you want to share for FAIR data principles, but others in your team don't find it that important) who has the authority to make the final decision? The first author? The supervisor/senior

author? This is just a hypothetical question but interested to know in case this comes up. (This is given the data sharing is allowed in your consent form.)”

In situations where there is disagreement within a research team about data sharing, and the consent form allows for it, the authority to make the final decision typically depends on a combination of research policies, the team’s roles, and any agreements made prior to the research project. To start with, I think you need to check the institutional policy as many universities and research organisations have clear policies on data sharing that researchers must follow. If no policy exists, the decision may fall to the principal investigator or supervisor, especially if they are the grant holder. However, if the issue remains unresolved, it would be wise to bring it to your institution’s research office or ethics committee for guidance.

“How do you ensure that using a pseudonymisation cannot identify an individual? are there standard tests?”

ICO has provided a detailed guidance on the following link about this https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/anonymisation/pseudonymisation/. ICO also provides information about Motivated Intruder tests for anonymised data https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/anonymisation/how-do-we-ensure-anonymisation-is-effective/#anonymisationprocessachieve}#motivatedintruder

“Why is ethnicity data given more weightage compared to personal data?”

Ethnicity data is classified as “special category data” under the UK GDPR. This means it is subject to stricter conditions for processing because it is considered more sensitive. The law recognizes that misuse or unauthorized disclosure of such data could lead to serious harms, including discrimination, stigmatization, or profiling. Therefore, additional safeguards are required to ensure its protection.

“Should there be a privacy notice in place when undertaking a research project or is providing information on data use/sharing in participant information sheet (PIS) and informed consent form (ICF) sufficient?”

It is possible to include the necessary privacy information in the Participant Information Sheet or Informed Consent Form, but it is often recommended to provide a separate privacy notice for clarity and compliance with the UK GDPR.

The privacy notice should include broader data protection information, such as the lawful basis for processing, how long data will be stored, individuals’ rights (e.g., right to access or delete their data), and contact details for the data controller or Data Protection Officer (DPO). This information is typically more detailed than what is covered in the information sheet or consent form, and the notice must be clear, easily accessible, and provided in writing.

You can reference the privacy notice in the information sheet or consent form, guiding participants to the full document, but it is still a good practice to provide it separately to ensure all data protection obligations are met.

“Can you expand on the implications of the UK GDPR / DPA for the open sharing of fully synthetic data? Particularly around necessary measures to establish the anonymity of respondents who contributed the original data on which synthesis was based.”

Although synthetic data is artificially generated, if it is derived from real personal data, it may still fall within the scope of data protection law if individuals can be identified, either directly or indirectly. The legal threshold for exemption is that the data must no longer relate to an identifiable individual, which is assessed using a contextual, risk-based approach: would re-identification be reasonably likely, considering all means likely to be used by someone?

Organisations must demonstrate that the synthesis process breaks any link to the individuals in the original dataset and that there is no meaningful risk of re-identification, even when considering external data sources. This involves testing for privacy leakage (such as membership inference or data memorisation), using appropriate privacy-preserving techniques (e.g. differential privacy or careful statistical sampling), and documenting the anonymisation process thoroughly. If these steps are met and the risk of re-identification is negligible, the synthetic data may be considered anonymous and shared openly without falling under the UK GDPR or DPA restrictions.