ADR UK response to ICO call for views: anonymisation, pseudonymisation and privacy enhancing technologies guidance November 2021

About ADR UK

ADR UK (Administrative Data Research UK) is an Economic and Social Research Council (ESRC) programme supported until September 2026. The partnership is supported by £90 million reinvestment from the ESRC, part of UK Research and Innovation (UKRI).

Its objective is to transform the way researchers access the UK's wealth of public sector data, to enable better informed policy decisions that improve lives. By linking together administrative data held by different parts of government and making it available for researchers, we are enabling vital research that has the potential to lead to more effective public services, in areas from improving education and healthcare to tackling crime.

ADR UK is made up of four national partnerships comprising ADR England, ADR Northern Ireland, ADR Scotland, ADR Wales, and the Office for National Statistics, coordinated by a UK-wide Strategic Hub. Each partnership is made up of academic and government partners, as well as dedicated secure services through which approved researchers can access de-identified administrative data. To find out more, visit the ADR UK website.

The below forms our response to the ICO call for views: anonymisation, pseudonymisation and privacy enhancing technologies guidance (November 2021). This accounts for the interests of all ADR UK partners, including those within each of the devolved nations – ADR England, ADR Northern Ireland, ADR Scotland and ADR Wales – and on a UK-wide level.

General comments

We welcome the ICO engaging with this agenda as it is of critical importance to the work of ADR UK. This guidance is addressing complex and evolving issues. It would be helpful at the outset to recognise this complexity and the confusion this causes for many organisations and researchers. We think part of the difficulty with this is that it is aiming to clarify an area that is still infused with grey areas.

Case studies

Cases continue to emerge on the parameters of 'anonymised' versus personal data. However, UK case law is not always consistent, is at odds with some of the guidance, and mostly predates the GDPR. Therefore, the publication of clear case studies designed to bring clarity around areas of legal uncertainty would be welcome. This would facilitate a practical approach to this complex area to be adopted, based on a broad understanding of the issues.
If this also recognised the 'usability' and 'utility' implications for organisations, this would also be extremely welcome. Ideally, these would recognise the importance of de-identifying data whilst still maintaining its usefulness for specific purposes. Also, of being able to effectively assess the relevant risks, so it was possible for organisations to navigate these issues with some degree of confidence.

Context

Some explanation about the broader context would also be helpful i.e., that there are both technical and legal issues to consider and both are evolving areas. This ICO guidance outlines that anonymisation is a good way forward that can reduce risk and help organisations share information fairly and proportionately. It highlights (in several places) that where data is effectively anonymised, the data protection regime no longer applies. However, it does not overtly acknowledge the complexity associated with anonymisation, the legal uncertainty when assessing the risk of re-identification, nor does it outline how this should be addressed by organisations.

In reality, there is no one generic answer because these issues are so context specific, meaning that whether anonymisation is appropriate or not will need to be considered on a case-by-case basis. The guidance does refer to the importance of context and the fact that anonymisation may not always be necessary, or possible. However, it needs to specifically draw out the factors that need to be taken into account when making this decision, so that organisations are better equipped to make this assessment.

Benefits and challenges

There are specific benefits and challenges associated with anonymisation and costs to consider too. When considering the benefits, the ESRC have funded the UK Data Service and others to produce and provide access to anonymised data for research for decades, most of which are useful for less advanced research, teaching etc., as per the data access spectrum. However, it must be acknowledged that data may be less useful once fully anonymised, so the benefits of being able to facilitate access to de-identified (rather than fully anonymised) data in a secure setting should also be acknowledged. Also, that organisations need to have sufficient skills, resources, technology and processes in place to make these decisions in a proportionate and appropriate manner. In summary, there is a balancing exercise to be done between

anonymisation, risk and competing interests and priorities. Further analysis and recognition of this dynamic would be helpful together with some examples of practical steps organisations can take.

ADR UK response

Anonymisation

On page 2 of the draft guidance, the ICO states "*Anonymisation offers an alternative way to use or share data by making sure that individuals are not identifiable*." We suggest that the ICO adds a reference to *accessing* data, as many secure trusted research environments (including all ADR UK ones) do not share data with researchers directly; researchers access the data via secure portal access only, in line with the 'five safes' principles. This mitigates many of the risks associated with the potential residual risk of the identifiability of the data, and is useful for data that really can't be usefully anonymised (as in administrative data and business microdata), although there are significant costs associated with maintaining these services. The ICO guidance should cover this, in addition to using or sharing data.

Pseudonymisation

In line with the guidance produced by the Medical Research Council (MRC) and ICO, and referenced in this guidance (page 3), pseudonymised information held in a safe environment such that the viewer (i.e., an accredited researcher) of the data is unlikely reasonable to be able to re-identify the records should be considered anonymised (i.e., non-personal) data.

This guidance sets out that pseudonymised data reduces the risk of reidentification but pseudonymised data is still personal data. We suggest that it be considered, as highlighted in the MRC and ICO guidance, to be on the continuum of identifiability and that pseudonymised data with the right controls in place could be considered *functionally anonymised.* (This is a concept attributable to Mark Elliot from the National Centre for Research Methods, which he describes in this video, Anonymisation: theory and practice; Mark Elliot (1/3)). For example, data held in a Digital Economy Act (DEA) - accredited trusted research environment with controls over use and user could be considered anonymous *in this context* even if outside the trusted research environment, it would be considered personal data.

The Northern Ireland Statistics and Research Agency (NISRA) Research Support Unit (RSU) (which facilitates researcher secure access to de-identified data) uses a dedicated secure network with staff who are fully trained and security cleared. NISRA has received accreditation for the RSU under the DEA. In order to meet the rigorous conditions for accreditation under the DEA, RSU have ensured that a set of detailed data management policies are in place and operationalised. This demonstrates compliance with the DEA. It would be useful for the ICO guidance and definitions to reflect these special circumstances.

The guidance describes 'de-identified' data as personal data that has undergone pseudonymisation. This is an accurate way of describing the data as it is handled by the data processor (e.g., ONS, NISRA) and ADR UK use this language to describe the administrative data research process. When it comes to how the data is used by the accredited researcher, it could be separately be described as anonymous if the ICO accept the argument about controls rendering it effectively anonymous to the user.

Deidentification

On page 4 of the draft guidance, the ICO states "*You should use this guidance if you are considering turning personal data into anonymous information.*" We suggest that the ICO adds a reference to

turning personal data into *de-identified* information, as it acknowledges it is extremely difficult to anonymise information (if it is physically possible) by carrying out fuzzy matching on the data while still retaining value and meaning to statisticians and researchers. If de-identified information is accessed via a trusted research environment that is accredited under a suitable process, for example under the DEA, and as such meets all the required security standards, then the de-identified data can be considered functionally anonymised, through a combination of the actions taken to create the de-identified dataset, and the environment in which the dataset is accessed (where it is not possible to deploy fuzzy matching techniques to link to other datasets).

If the ICO does include a reference to de-identified information, this should be done in a way that makes it clear what the differences are between the various terms, so people do not perceive that stripping out direct IDs (de-identification) is sufficient to render data anonymous. For this reason, there is a need for clear and consistent language (both among technicians and services – leaving aside language aimed at the public), which we discuss in the next section, 'Terms, definitions, and diagrams'.

On page 4 of the draft guidance, the ICO states "Anonymisation can help you to mitigate these risks and share information fairly and proportionately." There should be a complementary paragraph explaining how de-identification can help mitigate risks and access data fairly and proportionately. Further on, there should be sections of the guidance that explain how implementing de-identification can help, and what the benefits are. Also, how the term 'pseudonymisation' fits with de-identification and anonymisation, e.g., it depends on the context of how the data is accessed.

Terms, definitions, and diagrams

We think it would be useful for ADR UK to consider how we might want to use these definitions to further explain what it is that we do, as it is very important that researchers or others understand the distinction between these terms.

The definitions of terms such as 'anonymous information', 'anonymisation' and 'pseudonymisation' are difficult to follow mainly because definitions from different areas of law are included (Data Protection Act and UK GDPR). While this is helpful for the sake of completeness, it would be beneficial to explain the legal landscape a bit more fully so that it is clear why these different definitions exist and how organisations should manage the diverging definitions of key terms. This relates to a wider problem; that there are different definitions and interpretations of these terms and this is a by-product of the way the law has developed in this area.

The guidance highlights the importance of context, assessing risk and the need to exercise judgement, but more detail needs to be included to signpost and identify what factors organisations need to take into account when making judgements. In particular, more detail and specific guidance is needed in relation to the following:

On page 9 of the draft guidance, the ICO states, "*In the ICO's view, the same information can be personal data to one organisation, but anonymous information in the hands of another organisation. Its status depends greatly on its circumstances, both from your perspective and in the context of its disclosure*."

What factors does an organisation consider here to make an assessment?

On page 10 of the draft guidance, the ICO states, *"It is important to note that you must carefully assess each case individually based on the specific circumstances. This will help you to decide the effectiveness of an anonymisation technique and therefore whether the data is effectively rendered anonymous. Clearly, 100% or 'absolute' anonymisation is the most desirable position. At the same time, you will not always be able to state that a specific technique or set of controls will achieve these aims, particularly as technology changes over time. This means that even where you use anonymisation techniques, a level of inherent identification risk may still exist. However, this residual risk does not mean that particular technique is ineffective."*

We welcome the acknowledgement that information can be considered effectively anonymised if steps have been taken to ensure the risk of re-identification is sufficiently remote, even if such a risk has not been completely eliminated. What are the relevant factors to weigh up here? How do organisations assess the risk of re-identification and whether it is sufficiently remote? What role does the context in which the data are accessed play (for example, portal access via a DEA-accredited data processing environment, versus directly accessing data)? If absolute anonymisation renders the data no longer useful in a research context, then there is a balance to strike between data utility and risk of re-identification. It would be worth linking in references to GDPR, specifically the articles that allow use of special categories of personal data for scientific research.

On page 16 of the draft guidance, the ICO states *"If there are reasonably available means that could be used to re-identify individuals, then the data in question is not effectively anonymised. However, it is also important to consider the processing's context. For example, whether a dataset that is pseudonymised from your perspective has the same status from the perspective of another organisation you share it with."*

What are 'reasonably available means'? What aspects of the context require analysis?

The ICO report or indeed our translation of the ICO report for an ADR UK audience (once it is ready) would benefit from a diagram. Pictures can speak many words. It would also be helpful to include more examples and particular case studies that draw out key points and the difficulties that exist.

Relationship with other guidance

It would be helpful to understand how this relates to other guidance and relevant workstreams in this area. There is sector specific guidance (referred to by the ICO) as well as historic ICO guidance. How does this guidance relate to, update and modify other guidance?

In particular, there is an overlap between these issues and the research power in the Digital Economy Act 2017 (DEA) (that sets out a process for de-identifying personal information so it can be shared for research purposes). How does this guidance interrelate with this process? Some clarity on this would be helpful.