Unlocking the value of data:
Navigating anonymisation, pseudonymisation & PETs

02

Charlotte Halford
Partner
chalford@dacbeachcroft.com

Pete Given
Partner
pgiven@dacbeachcroft.com

In 2006, British mathematician Clive Humby stated that ***"data is the new oil"***. Twenty years later and this statement holds truer than ever. Organisations sit on a goldmine of personal data; from customer interactions and HR records to usage data from Internet of Things devices and digital platforms, personal data is at the core of how an organisation develops, markets and sells its products and services, and operates its business.

"Personal data powers our economy, from retail to hospitality to healthcare. Unlocking the potential of this data is key to encouraging economic growth and investment – as long as the public can trust it will be appropriately protected."

UK Information Commissioner, March 2025

There is an intense pressure to utilise and monetise data, driven largely by a desire to:

- **Have a competitive advantage:** data-driven insights and predictive analysis fuel smarter and faster decisions.
- **Implement operational efficiencies:** data can be used to streamline and automate processes.
- **Support AI adoption and training:** large, diverse datasets are essential for developing and refining AI models, enabling organisations to innovate responsibly and competitively.
- **Develop new products and services:** analysing data can lead to more desirable and personalised offerings.
- **Comply with regulatory incentives and mandates:** data-sharing mandates are increasingly being legislated for in sectors such as health, energy and finance.
- **Conduct evidence-based decision-making:** insights from analytical data drive stronger risk management, benchmarking and reporting.

However, the benefits of data utility come with an inherent challenge: data privacy compliance.

Anonymisation: unlocking value safely

Anonymisation is a method by which an organisation can harness the potential of data, but in a privacy-friendly way.

What is anonymisation?

Anonymisation is the process of turning personal data into anonymous information so that the individual is no longer identifiable. This limits the risk to individuals and can allow organisations to share the information more freely with other organisations or the public.

Truly anonymised data falls outside the scope of data protection laws. Consequently, it is important to ensure that where an organisation intends to rely upon anonymisation, it is correctly anonymising the data.

Organisations should consider the nature of the data; the purpose(s) it is used for; and the surrounding context. Key indicators of whether information is personal data include 'singling out' and 'linkability', as further examined in this piece.

What does the ICO say about anonymisation?

According to new anonymisation guidance issued by the Information Commissioner's Office (**ICO**), data is only considered anonymised if no individual can be identified by any party using means that are reasonably likely to be employed.

This test is contextual and risk-based, focusing on:

- **Means of reidentification:** The anonymisation must account for *"all the means reasonably likely to be used"* by any party, including the controller.
- **Nature of controls:** Technical safeguards such as access controls (e.g., firewalls, user permissions) may mitigate risk but do not constitute anonymisation if the controller retains the ability to reidentify individuals.
- **Context and role of the data recipient:** The likelihood of reidentification is assessed in light of the recipient's role, knowledge, and access to other data or tools.

The ICO's guidance sets out the two primary anonymisation techniques:

- **Generalisation:** which reduces the specificity of the data. This changes information that may identify someone so that it relates to multiple people. This means members of that group can't be identified or are no longer identifiable (for example, listing age ranges rather specific ages).
- **Randomisation:** which can be used to reduce the certainty that a record relates to a particular person. This changes information that may identify someone so that it cannot be definitively attributed to one person (for example, noise injection).

In addition, the ICO references the following techniques as supporting the above anonymisation techniques, but which will not themselves constitute anonymisation:

- Masking can reduce identifiability by deleting or supressing certain values or data records. While masking can be effective when used alongside generalisation and suppression, it is not considered an anonymisation technique on its own.
- Suppression is a disclosure control process where parts of the data are made unavailable to the user. The term is usually used to describe approaches like cell suppression, the removal of outliers and local suppression of particular values within microdata records.

The ICO emphasises that identifiability is not binary but exists on a spectrum.

> *At one end, data is fully identifiable; at the other, it is effectively anonymised. Between these extremes lies a grey area where re-identification risk depends on context, available resources, and technical measures.*

As such, when determining whether data is anonymous, organisations must assess the risk of:

- **Singling Out:** Can an individual be isolated within a dataset?
- **Linkability:** Could separate datasets be combined to reveal identity?
- **Inference Risk:** Could additional attributes allow identification indirectly?

This spectrum approach means anonymisation is rarely absolute and the risk of identification must be evaluated on a case by case basis, considering who holds the data and what auxiliary information they might access.

To operationalise this assessment, the ICO proposes the 'Motivated Intruder Test'.

> *ICO's "Motivated Intruder" test: Could a reasonably competent person, who has access to publicly available resources, use investigative practices to re-identify individuals?*

If the answer to the above question is yes, the data is not truly anonymised. This test reflects real-world conditions, where attackers may exploit open-source data, social media, or public records. It's a practical benchmark for organisations to gauge whether anonymisation techniques withstand plausible attempts at re-identification.

What does the case law say?

Two landmark cases illustrate the evolving interpretation of 'personal data' and identifiability under EU law:

- **SRB v EDPS (CJEU, 2025):** This case considered whether pseudonymised comments shared by the Single Resolution Board with Deloitte were 'personal data.' The Court clarified that pseudonymised data is not automatically personal data for every recipient; identifiability must be assessed contextually and from the recipient's perspective, considering whether the recipient has *"reasonable means"* to re-identify individuals. This is a pragmatic approach from the CJEU which requires assessment on a case-by-case basis. If organisations can evidence genuinely minimal re-identification risks, a contextual approach to determining whether pseudonymised data is personal data could enhance innovation.
- **Breyer v Germany (CJEU, 2016):** This case considered whether dynamic IP addresses could constitute personal data for a website operator if the operator has legal means to obtain additional information from an internet service provider to identify the user. The Court considered the *"means reasonably likely to be used"* test and concluded that they could. The test applied is the same concept that underpins the ICO's anonymisation guidance. It highlights that identifiability is not absolute but depends on practical and legal access to linking information.

Navigating the legal and contractual hurdles for anonymisation

Notwithstanding the seeming pragmatism of regulators and the courts, in the increasingly digital and data-driven world in which we live, effective anonymisation remains difficult to achieve in practice.

Where data that appears effectively 'anonymised' can be linked back to the underlying data subjects by combining it with other information which could enable reidentification, it will not meet the high standard for anonymisation required to fall outside data protection law. This might be the case where, for example, the organisation retains supplementary information, such as the original dataset or auxiliary data, or has access to additional data, such as a publicly available dataset. This principle was reinforced in *Breyer v Germany,* where dynamic IP addresses were deemed personal data because the website operator could legally access additional information to identify users. In practice, if an organisation retains linkage keys or supplementary datasets, the data remains pseudonymised and therefore subject to data protection obligations.

As highlighted in *SRB v EDPS*, identifiability is contextual: data may be personal in one organisation's hands but effectively anonymous in another's. This opens strategic opportunities. For example, creating a separate, dedicated entity for analytics, operating on segregated systems and dedicated employees without access to original identifiers, can transform pseudonymised data into anonymous data in that entity's hands, potentially even within the same group. This approach mitigates regulatory risk and unlocks data utility.

However, even when organisations have the technical capability to anonymise or pseudonymise data, they often face legal and contractual constraints. It is worth noting that the act of anonymisation itself, even if it produces effectively anonymised data, is an act of processing, which an entity may not be entitled to do, whether under contract or from a regulatory perspective.

A common scenario arises where an entity gains access to a dataset under a client contract that strictly limits its use to delivering services. These agreements may prohibit creating derivative datasets for internal purposes. Similarly, if the entity acts as a processor, it cannot lawfully repurpose the data without meeting controller obligations, such as establishing a lawful basis or complying with transparency requirements. This dual challenge means that, in practice, organisations may be unable to leverage data for innovation despite having the tools to anonymise it.

One solution is to address these hurdles at the contracting stage. Organisations can negotiate provisions that allow the creation of a dataset which, once the original data is no longer accessible (for example, at the end of the engagement), would be considered anonymous and therefore fall outside data protection law. This dataset could then be used for the organisation's own purposes, such as analytics, service optimisation, or product development. While residual issues remain, such as ensuring the anonymisation is robust and overcoming reluctance from data providers, this approach offers a pragmatic pathway to unlock data utility without breaching regulatory or contractual obligations.

Beyond anonymisation: strategic pathways to unlock data value

Even where full anonymisation proves elusive, organisations are not without options; there are still ways to extract value from the data they hold.

Pseudonymisation: a bridge between privacy and utility

When full anonymisation is not achievable, pseudonymisation offers a pragmatic alternative. It involves replacing or transforming identifiers so that data cannot be attributed to a specific individual without additional information, which is typically held separately.

Common techniques include hashing, encryption, and tokenisation. Unlike anonymisation, pseudonymisation preserves much of the dataset's richness, enabling advanced analytics, AI training, and cross-functional insights while reducing direct identifiability.

However, pseudonymisation is not a silver bullet. Data protection laws still apply because the possibility of re-identification remains if linkage keys or auxiliary datasets exist. Organisations must assess the robustness of their approach against realistic attack scenarios, such as exhaustive searches or dictionary attacks. The ICO's guidance on pseudonymisation stresses that pseudonymisation should not be treated as anonymisation; it is a risk-reduction measure, not an exemption from compliance.

Despite these constraints, pseudonymisation can unlock significant value. It enables organisations to share data internally or with partners under controlled conditions, supporting innovation without exposing raw identifiers. For example, pseudonymised datasets can power predictive models, benchmarking, and service optimisation while maintaining a privacy-conscious posture. Combined with strong governance, such as segregating linkage keys, implementing contractual safeguards, and conducting Data Protection Impact Assessments (**DPIAs**), pseudonymisation becomes a strategic enabler.

In a regulatory environment increasingly focused on 'privacy by design,' organisations that master pseudonymisation can position themselves ahead of competitors, balancing compliance with data-driven growth.

Legislative reform: a new hope for data utility?

If anonymisation and pseudonymisation alone are not the panacea, could legislative reform offer a lifeline for organisations seeking to unlock data value? Recent policy developments suggest that lawmakers are beginning to recognise the tension between privacy protection and innovation and are taking steps to recalibrate the balance.

The UK's Data (Use and Access) Act 2025 (**DUAA**) is a prime example. By introducing broad consent provisions for scientific and statistical research, it reduces friction for organisations that want to share and analyse data for socially beneficial purposes. The DUAA also streamlines compliance by relaxing certain privacy notice obligations, signalling a shift toward enabling responsible data use rather than simply restricting it. For organisations operating in research-heavy sectors - healthcare, energy and financial services - this could be transformative, provided they embed robust governance to maintain trust.

Across the Channel, the EU's Digital Omnibus Package, published in November 2025, proposes amendments to the EU GDPR that could redefine the status of pseudonymised data. Under the draft, pseudonymised datasets may, in certain contexts, fall outside the definition of personal data, potentially unlocking new opportunities for cross-border collaboration and AI development. While implementation remains uncertain, the direction of travel is clear: regulators are exploring pragmatic solutions to reconcile privacy with innovation.

> *"Pseudonymisation is a way of reducing risk and improving security. It is not a way of transforming personal data to the extent the law no longer applies."*
>
> ICO Guidance on Pseudonymisation

Privacy-enhancing technologies: unlocking value without compromise

When anonymisation and pseudonymisation reach their limits, organisations are increasingly turning to Privacy-Enhancing Technologies (**PETs**) as a way to reconcile two competing priorities: extracting insight from data and safeguarding individual privacy. PETs represent a shift from traditional compliance tools to advanced technical solutions that enable collaboration, analytics, and AI development without exposing raw personal data. These technologies are not just about compliance, they are about embedding trust into data strategies and enabling responsible growth in a data-driven economy.

PETs encompass a range of techniques designed to minimise personal data use, maximise information security to preserve privacy, and empower individuals. According to the ICO's guidance, PETs can:

- **Reduce the identifiability of individuals:** through synthetic data generation and differential privacy.
- **Hide or shield information:** using zero-knowledge proofs, homomorphic encryption and trusted execution environments.
- **Split datasets for secure computation:** through secure multi-party computation and federated learning.

The benefits are significant. PETs help organisations demonstrate data protection by design and default; comply with data minimisation principles and maintain robust security whilst enabling analytics and AI training. They can also support anonymisation or pseudonymisation strategies, control access to sensitive datasets; and reduce breach risks. For AI-driven initiatives, PETs are particularly valuable, allowing models to learn from distributed or sensitive data without compromising privacy.

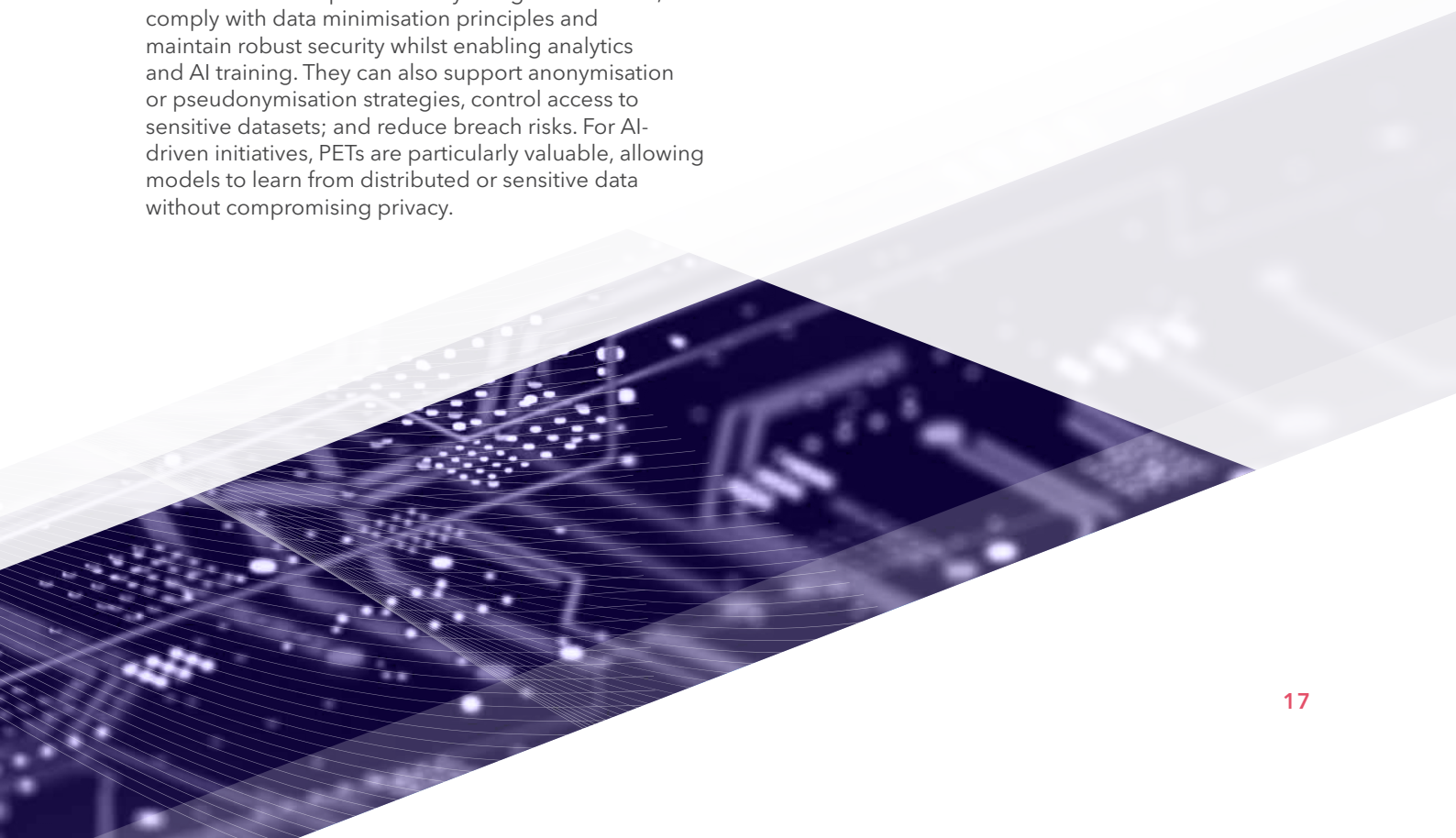
However, PETs are not without limitations. Most PETs still involve processing personal data, meaning organisations must ensure lawful, fair, and transparent handling.

In addition, they face practical challenges such as scalability, lack of mature standards, vulnerability to sophisticated attacks, and implementation errors. Expertise gaps and insufficient organisational measures can undermine effectiveness.

To mitigate these risks, organisations should conduct DPIAs and integrate PETs into broader governance frameworks. The strategic question is not whether PETs will become mainstream, but how quickly organisations can adopt them to turn privacy into a competitive advantage.

The ICO's guidance is clear: *"the purpose of many PETs is to enhance privacy and protect the personal data you process, rather than to anonymise that data. This means that:*

- many PET use-cases still involve personal data; and
- when you deploy such techniques, you still need to meet your data protection obligations".



Synthetic data

Synthetic data is a particular PET that we envisage will see increased uptake with the rise in creation and deployment of AI tools. Synthetic data is 'artificial' data generated by data synthesis algorithms. There are two types of synthetic data:

- **'Partially' synthesised data:** where only some variables of the original data are synthesised (for example, just synthesising location in an A&E admission dataset).
- **'Fully' synthesised data:** where all variables of the original dataset are synthesised (for example, synthesising name, location, admission time and reason for admission in an A&E admission dataset).

Synthetic data is a useful tool for training artificial intelligence models in environments where real data is scarce or sensitive. For example, Mastercard found that using synthetic data gave them a competitive advantage when performing analysis in emerging markets. Synthetic data benefits from preserving statistical integrity, eliminating directly identifiable data, and being cost-efficient. The ICO advises that organisations should consider using synthetic data as a tool for generating non-personal data in situations where it does not need to, or cannot, share personal data.

The challenge for organisations is whether the synthetic data used is an accurate substitute for the original data. Quality controls should be put in place to ensure that poorly generated synthetic data does not skew models and datasets (for example, through biases).

Practical measures for organisations to unlock data value

Turning data into a strategic asset requires more than technical capability, it demands foresight, governance, and contractual clarity. To unlock value effectively and compliantly, organisations should:

- **Define a data vision:** Go beyond short-term needs. Map current and future data requirements, identify sources, and assess whether internal or external datasets can deliver competitive advantage. This is the foundation for any data-driven strategy.
- **Embed governance by design:** Implement robust frameworks that integrate privacy, security, and ethical considerations from the outset. Use DPIAs not as a tick-box exercise, but as a strategic tool to evaluate risk and opportunity.
- **Master advanced techniques:** Understand the spectrum of options, anonymisation, pseudonymisation, and PETs, and when each is appropriate. Keep pace with evolving case law and legislative reforms that may redefine what counts as personal data.
- **Engineer contractual flexibility:** Contracts often dictate what you can and cannot do with data. Negotiate provisions that allow for anonymisation or pseudonymisation for analytics and innovation and consider future-proofing agreements to accommodate emerging technologies and regulatory changes.
- **Invest in capability and culture:** Technology alone won't deliver value. Build internal expertise in data ethics, privacy engineering, and AI governance. Foster a culture where compliance is seen not as a constraint but as a catalyst for trust and innovation.

Concluding thoughts

Unlocking the value of data requires careful consideration of the legal, regulatory and contractual controls that may attach to that data. Anonymisation, pseudonymisation and PETs can all be used to unlock the value of data in a manner that complies with those controls, but there is no one size fits all approach and each has its advantages and disadvantages.

Organisations that adopt robust governance and contractual safeguards will be best positioned to harness data responsibly and competitively in an evolving regulatory landscape.

Data Utility v Privacy Risk