Not All Personal Data Can Be Anonymised — But All Personal Data Must Be Governed

By: Jaiz Anuar

10 July 2026 · Data Governance · Privacy · Security Architecture · 10 min read

Anonymisation is not only a privacy control. It is a governance and architecture decision about when personal data truly needs to remain identifiable.

“Just anonymise the data.”

It sounds simple. It sounds responsible. It sounds like the right answer in any privacy or data protection discussion.

But in real enterprise environments, anonymisation is rarely that simple.

Personal data does not sit quietly inside one application. It moves. It is copied. It is reported. It is extracted. It is shared. It is archived. It appears in dashboards, data warehouses, operational reports, audit trails, vendor support files, test environments, APIs, and sometimes even spreadsheets created outside formal system controls.

This is why anonymisation cannot be treated as a simple technical button.

It is not only a privacy control. It is a business, architecture, governance, and risk decision.

The difficult truth is this: not all personal data can be anonymised.

Some data must remain identifiable because the organisation still needs to know who the customer is, who owns the account, who performed the transaction, who approved the request, who raised the complaint, or who is entitled to receive a service.

A bank cannot fully anonymise customer data that is still required for account servicing. An investment institution cannot anonymise investor records that are still needed for regulatory reporting, dispute handling, audit, tax, or legal retention. A human resource system cannot anonymise employee records that are still required for payroll, benefits, disciplinary records, or statutory obligations.

When data is truly anonymised, it should no longer be possible to identify the individual, directly or indirectly. That is the point—but it is also the challenge.

If the business still needs to re-identify the individual, then full anonymisation may not be appropriate for that specific use case.

Ask the Better Question

This is where many organisations make the mistake. They ask: “Can we anonymise this application?”

But the better question is: Where does the personal data go, who uses it, why do they need it, and does it still need to be identifiable at that point?

That difference matters.

Anonymisation may not be feasible in the core production system because the business still requires identifiable data. But the same conclusion may not apply to downstream reporting, analytics, testing, development, research, external sharing, data lakes, archival, or vendor support activities.

A production system may need identifiable customer records. But does a management dashboard need names and identification numbers? A regulatory report may require traceability. But does every internal operational report require full personal identifiers?

A development team may need realistic data to test system behaviour. But does that mean it needs real customer data? A vendor may need data to troubleshoot an incident. But does that mean raw production data should be shared without masking, tokenisation, or controlled access?

This is why the conversation must move beyond anonymisation alone.

If full anonymisation is not feasible, that should not become permission to do nothing. It should trigger a deeper control discussion.

Reduce Exposure Where Identifiability Is Not Needed

There are practical ways to reduce exposure without forcing full anonymisation everywhere:

Mask data at the user-interface layer.

Tokenise identifiers before they are used for analytics.

Use aggregated data instead of individual-level data in reports.

Replace production data with synthetic data for testing.

Restrict and encrypt archived data, making it available only through approved workflows.

Prevent APIs from exposing unnecessary personal-data fields.

Monitor data extracts and document data lineage.

Limit re-identification access to a small group of authorised users.

These are not just privacy questions. They are security architecture questions.

From an architecture perspective, personal data must be governed across the full data lifecycle: the source application, database, integration layer, API, ETL pipeline, reporting platform, data warehouse, data lake, archive, backup, non-production environment, third-party connection, and manual export process.

A system can appear compliant at the application level while still leaking personal data through downstream channels. This is often where the real risk sits.

The source system may have proper access control, but the same data may later appear in an Excel extract with weaker protection. The database may be encrypted, but a report may expose full identifiers to too many users. The application may mask sensitive fields, but the API may still return raw values. Production may be well controlled, but test environments may contain copied personal data with broader access.

The organisation may have a data-retention policy, but no one may know whether retained data still needs to remain fully identifiable for the entire retention period.

Retention Is Not the Same as Identifiability

“Data retention” and “identifiability” should not be treated as the same thing.

A regulation may require a record to be retained. But that does not always mean every data element must remain visible to everyone throughout the entire retention period.

There may be a need to retain the record, but access can still be restricted. Certain fields can be masked. Some datasets can be pseudonymised. Some reports can be aggregated. Some older data can be moved into a controlled archive. Some secondary use cases can rely on anonymised or synthetic data.

The point is not to force anonymisation everywhere. The point is to ensure identifiability is justified.

Personal data should remain identifiable only where there is a valid business, legal, regulatory, operational, or audit reason. Where that reason does not exist, the organisation should reduce exposure.

Governance Decides What Tools Cannot

This is where governance becomes more important than the tool.

Tools can mask, tokenise, encrypt, discover, classify, and monitor data. But tools cannot decide the business purpose. They cannot define retention justification, approve data sharing, determine whether a report truly needs personal identifiers, or replace accountability.

Governance must answer the harder questions:

Who owns the data, and who is allowed to use it?

For what purpose, for how long, and at what level of detail?

Can it be shared externally, used for testing or analytics, copied into a data lake, or used to train AI models?

Can it be re-identified, who can approve that action, and what evidence supports the decision?

Without these answers, anonymisation becomes a checkbox. Pseudonymisation becomes a vague recommendation. Masking becomes inconsistent. Tokenisation becomes application-specific. Data sharing becomes subjective. And privacy by design becomes a slogan rather than a working control.

Make the Decision Part of the Operating Model

The better approach is to embed anonymisation and pseudonymisation decisions into existing governance processes: solution architecture review, privacy impact assessment, data governance, secure software development lifecycle, cloud and SaaS onboarding, third-party risk review, API governance, data-sharing approval, reporting design, and non-production data management.

Each new use case involving personal data should ask a few basic questions:

Is identifiable data required, or can the purpose be achieved with less data?

Can direct identifiers be removed, and can indirect identifiers still re-identify the person?

Can the data be aggregated, masked, or tokenised to reduce exposure?

Is re-identification required, who can perform it, and how will it be logged and monitored?

When should the decision be reviewed?

This is how privacy control becomes operational: not by saying every application must anonymise everything, and not by saying anonymisation is impossible and moving on.

It happens by making every personal-data use case explain why the data needs to remain identifiable.

Final Thought

The goal is not to anonymise everything.

The goal is to ensure personal data is only identifiable when it has a valid reason to be identifiable.

Because not all personal data can be anonymised. But all personal data must be governed.