DATA MASKING PROCEDURE

Document Reference: QMS-SEC-PRO-001

Version: 1.0

Classification: INTERNAL ENGINEERING ONLY

Role	Name & Signature	Position	Date
Prepared by			
Reviewed by			
Approved by			

Design notes and regulatory alignment: https://github.com/Lewiskunta/data-masking-procedure

TABLE OF CONTENTS

1. AMENDMENT RECORD

Reviewed at minimum annually or following a significant data exposure incident, regulatory change, or material change to the data estate. All revisions are subject to the same approval requirements as the original document.

Date	Issue	Rev	Page	Subject	Revised By	Approved By

2. GENERAL

2.1 Scope

This procedure applies to all personally identifiable information (PII), authentication credentials, financial data, health data, and other sensitive or regulated data processed, stored, or transmitted by the organisation. It governs the selection, implementation, testing, and audit of data masking controls across:

- Production database exports used in non-production environments
- Data shared with third parties, contractors, or external auditors
- Data used in development, QA, staging, and training environments
- Analytical and reporting pipelines that process production data
- Log files, monitoring outputs, and observability data
- API responses that may include sensitive field values
- Backup and disaster recovery data sets
- Support tooling that exposes customer data to internal staff

This procedure is technology-agnostic. It applies regardless of whether the underlying platform is cloud-native (IaaS, PaaS, SaaS), on-premises, containerised, or hybrid. Platform-specific implementation notes are provided in Section 4 where relevant.

Data masking applied at the point of collection or ingestion (i.e., data that is never stored in unmasked form) is out of scope for this procedure but must conform to the data classification requirements in Section 3.1.

2.2 Purpose

- Define the minimum masking controls required for each data classification level
- Establish approved masking techniques and their permitted use cases
- Specify implementation requirements for structured and unstructured data masking
- Define testing and validation gates before masked data is promoted to any shared environment
- Provide an auditable evidence trail sufficient for ISO 27001:2022, PCI DSS v4.0, GDPR, and applicable national data protection law compliance
- Govern the incident response process for masking failures and exposure events

2.3 Guiding Principles

- Mask by default: sensitive data is masked before it leaves the production boundary. Unmasking requires explicit authorisation and produces an audit record.
- Technique selection is determined by data classification and use case, not by convenience. The technique selection matrix in Section 4.2 is mandatory.

- Masking does not substitute for access control. Masked data environments still require role-based access controls. Masking reduces the blast radius of an access control failure; it does not eliminate the need for access control.
- Reversibility is a risk, not a feature. Reversible masking techniques (tokenisation, encryption) require secure key and token management. Loss of the mapping constitutes data loss. Compromise of the mapping constitutes a breach.
- Referential integrity must be preserved. Masking applied to a field that is referenced by a foreign key or join must produce a consistent substitute value across all tables in the dataset.
- Masked data must remain functionally usable for its intended purpose. Masking that breaks application functionality or test validity is a masking implementation defect.

2.4 Responsibility for Implementation

Role	Owns	Accountable Output
DPO	Procedure ownership; regulatory compliance; breach notification	Approval signature on Appendix A for all third-party and Regulated data sharing; re-identification risk assessments (Section 7.2); supervisory authority notification under Section 9.1
Information Security	Technique approval; exception review; pipeline configuration approval; audit programme	Signed exception forms (Appendix B); quarterly masking control audit report; key management audit report (Section 8.1)
DevOps / Platform Engineering	Masking pipeline build and maintenance; KMS service account configuration; synthetic dataset generation	Pipeline peer review records per Section 5.1.3; salt retrieval implementation per Section 5.4; synthetic datasets for developer use per Section 5.3.5
Database Administrators	Structured data masking execution; referential integrity verification	Completed pre-masking checklist (Appendix C); Gate G2 and G5 sign-off (Section 7.1
QA / Test Engineering	Masked dataset validation; red-team pattern detection	Gate G3 and G4 sign-off (Section 7.1); quarterly detection test results (Section 7.2)
All staff handling sensitive data	Classification compliance; masking request initiation	Completed Appendix A before any masking operation or data sharing event

2.5 References

- ISO 27001:2022, Annex A controls 8.11 (data masking) and 8.12 (data leakage prevention)
- ISO 27701:2019, Privacy Information Management Systems
- GDPR Article 25 (data protection by design and by default) and Article 32 (security of processing)
- PCI DSS v4.0, Requirements 3.3 (SAD protection), 3.4 (PAN rendering), and 3.5 (cryptographic key management)

- NIST SP 800-188, De-Identification of Government Datasets
- NIST SP 800-57, Recommendation for Key Management
- OWASP Data Masking Cheat Sheet
- Internal: Data Classification Policy
- Internal: Incident Response Procedure
- Internal: Configuration Management Procedure
- Internal: Access Control Policy

3. DATA CLASSIFICATION AND MASKING REQUIREMENTS

3.1 Classification Levels

All data processed by the organisation is assigned a classification level at the point of creation or ingestion. Classification determines the mandatory masking controls, the permitted environments for unmasked data, and the approved sharing conditions.

Level	Definition	Examples	Masking Requirement	Permitted Unmasked Environments
Public	Information with no sensitivity that can be disclosed without restriction	Published documentation, product descriptions, marketing copy	None required	All
Internal	Non-sensitive operational information restricted to employees and contractors	Internal communications, non-sensitive operational metrics, anonymised usage statistics	Masking required before sharing externally or with contractors lacking NDA	Production, staging (with access controls)
Confidential	Sensitive information whose unauthorised disclosure would cause material harm	Customer contact details, internal financial reports, employee records, non-payment business data	Masking required in all non-production environments and before any external sharing	Production only (with RBAC); staging with DPO approval
Restricted	Highly sensitive information whose unauthorised disclosure would cause severe harm or regulatory breach	Authentication credentials, session tokens, encryption keys, full PII datasets, health records	Masking required in all environments except production. Irreversible masking preferred unless application function requires reversibility	Production only (with strict RBAC and MFA); no exceptions without DPO approval

Regulated	Information subject to specific statutory or contractual data protection requirements	Payment card data (PCI DSS), health records (HIPAA where applicable), data of EU data subjects (GDPR)	Masking technique must satisfy the applicable regulatory standard. PCI DSS requires PAN truncation or tokenisation as a minimum. GDPR requires pseudonymisation or anonymisation for non-production use	Production only under full regulatory controls; non-production requires DPO and CISO sign-off

3.2 Data Types Requiring Masking

The following data types are subject to masking under this procedure. This list is not exhaustive. Data owners are responsible for identifying and classifying new data types as they are introduced.

Data Type	Examples	Classification	Mandatory Masking Technique(s)
Direct PII -- Identity	Full name, date of birth, national ID number, passport number, tax ID	Restricted	Pseudonymisation or substitution; format-preserving where application requires name format
Direct PII -- Contact	Email address, phone number, physical address, IP address	Confidential to Restricted	Substitution or partial masking; email domain may be preserved for format-dependent applications
Authentication data	Passwords, PINs, security question answers, MFA seeds	Restricted	Irreversible hashing (bcrypt, Argon2) in all environments; never stored or transmitted in cleartext
Session and access tokens	OAuth tokens, API keys, session IDs, JWTs	Restricted	Tokenisation or nullification; tokens must be invalidated before data export, not merely masked
Payment card data	PAN (Primary Account Number), CVV/CVC, expiry date, cardholder name	Regulated (PCI DSS)	PAN: tokenisation or truncation to last 4 digits (PCI DSS Req 3.4). CVV/CVC: never stored post-authorisation (PCI DSS Req 3.3.1). Full card data: format-preserving encryption if storage is required
Bank account data	Account number, sort code, IBAN, SWIFT/BIC	Regulated	Tokenisation or partial masking (last 4 digits visible maximum)
Financial transaction data	Transaction amounts, merchant codes, payment references	Confidential to Regulated	Pseudonymisation of customer-linked fields; transaction amounts may be preserved for

			analytics if customer linkage is removed
Health and medical data	Diagnoses, prescriptions, clinical notes, health insurance identifiers	Regulated	Pseudonymisation; direct identifiers removed; quasi-identifiers (age, postcode, gender) generalised to prevent re-identification
Biometric data	Fingerprints, facial recognition templates, voice prints	Regulated	Irreversible transformation or deletion; biometric templates must not be exported to non-production environments under any circumstances
Business-sensitive configuration	Database schemas, resource allocation models, proprietary algorithms, pricing models	Confidential	Obfuscation or structural scrambling; specific field values replaced with representative but non-real values
Log and telemetry data	Application logs, access logs, error traces, APM data	Confidential	PII and credential scrubbing before log persistence; regex-based pattern matching for known sensitive field formats
Backup and archive data	Full or incremental database backups, archived datasets	Matches source classification	Same masking requirements as source data apply before backup is exported to non-production storage

4. APPROVED MASKING TECHNIQUES

4.1 Technique Reference

The following techniques are approved for use under this procedure. Technique selection must follow the matrix in Section 4.2. Use of a technique not listed here requires written approval from the Information Security team and a documented rationale.

Technique	Mechanism	Typical Use Cases	Reversibility and Risk	Implementation and Compliance Notes
Tokenisation	Replaces a sensitive value with a non-sensitive surrogate token. A secure token vault maintains the mapping between token and original value.	Payment card PAN, account numbers, national ID numbers, customer IDs requiring referential integrity across systems	Reversible (requires token vault access). Appropriate when the original value must be recoverable by authorised parties. Token vault is itself a critical	PCI DSS: satisfies PAN de-scoping requirement when implemented correctly. Must use cryptographically random tokens with no mathematical relationship to the original value.

			security asset requiring HSM-grade protection.	
Format-Preserving Encryption (FPE)	Encrypts data such that the output has the same format, length, and character set as the input. Uses AES-FF1 or AES-FF3-1 (NIST SP 800-38G).	PAN, phone numbers, national ID numbers where downstream application requires field format to be maintained	Reversible (requires decryption key). Key management is critical. FPE keys must be stored in an HSM or equivalent key management service.	PCI DSS: satisfies PAN protection requirement. NIST-approved algorithms only (FF1, FF3-1). FF3 has known weaknesses against related-key attacks; FF1 is preferred for new implementations.
Pseudonymisation	Replaces direct identifiers with pseudonyms using a consistent mapping table stored separately from the data. The data can be re-identified by an authorised party with access to the mapping table.	PII datasets used for analytics, test data generation, GDPR-compliant data processing	Reversible (requires mapping table access). Satisfies GDPR Article 4(5) definition but does not constitute anonymisation. Data remains personal data under GDPR.	GDPR: pseudonymised data is still personal data. The pseudonymisation key/mapping must be stored separately from the pseudonymised dataset and must not be accessible to the data processor.
Data Substitution	Replaces sensitive values with realistic but entirely fabricated values drawn from a reference dataset or generated by a synthetic data library.	Names, addresses, email addresses, phone numbers in test and development environments	Irreversible (no mapping to original). Reference datasets must not contain real data. Generated substitutes must pass format validation for downstream application compatibility.	Preferred technique for development and QA environments where the original value is not needed. Substitution libraries (Faker, Mimesis) must be seeded to produce consistent values within a dataset for referential integrity.
Data Scrambling	Randomly shuffles character values within a field or shuffles values between rows within a column while preserving data type, length, and format.	Numeric fields, alphanumeric identifiers, free-text fields where statistical distribution must be preserved	Irreversible. Column-level shuffling preserves statistical distribution but may expose patterns if the dataset is small. Row-level shuffling breaks the association	Not suitable for fields with high cardinality and low dataset size (risk of re-identification). Not suitable for fields with functional dependencies across tables without consistent scrambling

			between a value and its original record owner.	applied across all related tables.
Partial Masking (Data Redaction)	Reveals only a defined portion of a sensitive value. The remainder is replaced with a masking character (typically an asterisk or X).	Display of PAN in UI (last 4 digits), display of account numbers in support interfaces, log file redaction	Irreversible. The visible portion must not be sufficient for re-identification or fraudulent use in isolation. The minimum visible portion must be defined per field type.	PCI DSS: display of last 4 digits of PAN is the maximum permitted for most use cases. First 6 and last 4 digits may be displayed only if there is a documented, legitimate business need. CVV/CVC must never be displayed in any form post-authorisation.
Nullification	Replaces sensitive field values with NULL or a defined placeholder value (e.g., REDACTED).	Log file sanitisation, removal of sensitive fields from API responses when the consumer does not require them, removal of deprecated PII fields during data retention enforcement	Irreversible. Nullification of a NOT NULL database column requires schema adjustment or a defined placeholder constant. Referential integrity constraints must be assessed before nullification is applied to foreign key columns.	Simplest technique with the lowest residual risk. Appropriate when the field value has no functional use in the target environment. Not appropriate when the field is required for application logic, test coverage, or regulatory audit purposes.
Generalisation	Replaces a specific value with a less specific but accurate range or category. Reduces precision rather than removing the value.	Age (replace exact age with range: 30-39), postcode (replace full postcode with sector), transaction amount (replace exact amount with band)	Irreversible. The generalisation intervals must be defined such that no individual can be singled out from the resulting dataset (k-anonymity principle: each combination of quasi-identifiers must appear in at least k records).	Used specifically for analytical and reporting datasets where statistical utility must be preserved. Requires a privacy engineer to define appropriate generalisation intervals. Generalisation alone is insufficient for datasets with a small number of records per group.

Irreversible Hashing	Applies a one-way cryptographic hash function to a sensitive value. The hash output is stored in place of the original.	Passwords, PINs, security answers (in all environments), de-duplication of PII without retaining original values	Irreversible. Must use a salted, slow hashing algorithm (bcrypt, Argon2id, scrypt) for passwords and credentials. SHA-256 or SHA-3 (without salt) is not acceptable for password storage. MD5 and SHA-1 are prohibited.	Passwords must never be stored or transmitted in any reversible form. Hashing is the only acceptable technique for authentication credential storage. For de-duplication purposes, a keyed HMAC may be used to maintain consistency without the reversal risk of tokenisation.
Data Synthesis (Synthetic Data Generation)	Generates a statistically representative dataset that contains no real records. The synthetic dataset shares the statistical properties of the source dataset but cannot be traced back to any real individual.	Large-scale test environments requiring production-like data volumes and distributions; ML model training; performance testing	Irreversible by definition. Requires a validated synthetic data generation tool and a quality assessment step to confirm the synthetic dataset's statistical fidelity to the source.	Emerging best practice for test data management in regulated industries. Does not constitute personal data under GDPR if implemented correctly (no record in the synthetic dataset corresponds to a real individual). Validation methodology must be documented.

4.2 Technique Selection Matrix

The matrix below defines the mandatory technique for each data type and target environment combination. Where multiple techniques are listed, the first is preferred. Deviations require a completed Data Masking Exception Request (Appendix B).

Data Type	Dev / QA	Staging	Analytics	Third-party Share	Support / Helpdesk
PAN / Payment Card	Tokenisation	Tokenisation	Tokenisation + Partial (last 4)	Tokenisation	Partial (last 4 digits)
CVV/CVC	Nullification	Nullification	Nullification	Nullification	Nullification
Bank account number	Tokenisation or Partial	Tokenisation	Partial (last 4)	Tokenisation	Partial (last 4)
Full name	Substitution	Substitution or Pseudonymisation	Generalisation or Pseudonymisation	Pseudonymisation	Partial or Substitution

Email address	Substitution	Pseudonymisation	Pseudonymisation	Pseudonymisation	Partial (domain visible)
Phone number	Substitution	Pseudonymisation	Generalisation	Pseudonymisation	Partial (last 3 digits)
Physical address	Substitution	Pseudonymisation	Generalisation (postcode sector only)	Pseudonymisation	Partial (postcode only)
National ID / Tax ID	Substitution	Pseudonymisation	Nullification	Tokenisation	Partial (last 4 chars)
IP address	Substitution	Pseudonymisation	Generalisation (first 2 octets)	Pseudonymisation	Partial (first 2 octets)
Password / PIN	Hashing (Argon2id)	Hashing (Argon2id)	Nullification	Nullification	Nullification
Session / API token	Nullification	Nullification	Nullification	Nullification	Nullification
Health / Medical data	Pseudonymisation + Generalisation	Pseudonymisation	Generalisation + Pseudonymisation	Pseudonymisation (DPO approval required)	Partial or Nullification
Biometric data	Prohibited from export	Prohibited from export	Prohibited from export	Prohibited from export	Prohibited from export
Log file PII	Scrubbing / Nullification	Scrubbing / Nullification	Scrubbing / Nullification	Scrubbing / Nullification	Scrubbing / Nullification

4.3 Prohibited Practices

- Storage or transmission of CVV/CVC, full magnetic stripe data, or PIN block data in any form after transaction authorisation
- Storage or transmission of passwords or authentication credentials in cleartext or using reversible encryption
- Export of biometric templates to any non-production environment or to any third party
- Use of MD5, SHA-1, DES, 3DES, or RC4 in any masking or encryption control under this procedure
- Use of unsalted hash functions for password or credential masking
- Masking that relies on security through obscurity (e.g., encoding data in base64 and treating it as masked)
- Manual ad-hoc masking without reference to this procedure and without logging the masking action
- Storing tokenisation or FPE keys in the same database or storage system as the data they protect

5. IMPLEMENTATION REQUIREMENTS

5.1 Structured Data Masking

Structured data masking applies to relational databases, data warehouses, and any tabular data store. The following steps are mandatory for all structured data masking operations.

5.1.1 Data Discovery and Classification

- Run automated data discovery tooling against the target database to identify columns containing sensitive data. Discovery must cover all schemas, not only the primary application schema.
- Classify all discovered sensitive columns against the classification levels in Section 3.1. The classification output is the input to the masking rule definition in Step 5.1.2.
- Document the discovery results in the Data Masking Request Form (Appendix A). The form must be approved by the data owner before masking proceeds.
- Discovery must be re-run whenever a schema migration adds or modifies columns. Schema migrations that add columns containing sensitive data require an updated masking rule set before the migrated schema can be used in a non-production environment.

5.1.2 Masking Rule Definition

- For each classified column, select a masking technique from the matrix in Section 4.2. Document the selected technique in the masking rule set.
- Identify all foreign key relationships and referential constraints involving masked columns. Masking rules must be applied consistently across all tables sharing a relationship. Inconsistent masking across related tables produces datasets that break application joins and invalidates test results.
- Where substitution is used, define the substitution reference dataset and confirm it contains no real data.
- Where tokenisation or FPE is used, confirm that key management is in place per Section 5.4 before masking is executed.

5.1.3 Masking Execution

- Masking is executed against a copy of the production dataset, not the live production database. No masking tool or script has write access to the production database.
- Masking execution is logged. The log records: timestamp, operator, source database reference, tables and columns processed, technique applied per column, and row counts before and after masking.
- Masking pipelines are version-controlled and subject to peer review per the Code Review Procedure (QMS-DEV-PRO-003). Masking scripts are treated as production code.
- Where automated masking pipelines are used for environment refresh, the pipeline must complete successfully and produce a validation report (Section 5.1.4) before the refreshed environment is made available to users.

5.1.4 Validation

- Validate that no original sensitive values are present in the masked dataset. For each masked column, run a sample check comparing masked output against the original. Sample size must be at least 5% of rows or 500 rows, whichever is greater.
- Validate referential integrity: all foreign key relationships must resolve correctly in the masked dataset.
- Validate that masked values conform to the expected format for the column (e.g., masked email addresses remain valid email format if the application requires it).
- Validate application functionality against the masked dataset in a representative test run before the environment is released. Test pass rate must meet the QA exit criteria defined in the SDLC Procedure (QMS-DEV-PRO-001).
- Automated pattern detection must scan the masked dataset for any residual sensitive data patterns (credit card numbers, national ID formats, email patterns). Any match is a masking failure and requires the masking pipeline to be corrected and re-executed before the dataset is used.

5.2 Unstructured Data Masking

Unstructured data masking applies to documents, emails, log files, object storage contents, API payloads, and any non-tabular data containing sensitive information.

5.2.1 Content Discovery

- Use content inspection tooling to scan unstructured data repositories for sensitive content. Scanning must cover file contents, not only filenames.
- Pattern libraries for discovery must include: email address patterns, phone number patterns (including country-specific formats), national ID number formats for all applicable jurisdictions, credit card number patterns (Luhn-validated), IP address patterns, and any organisation-specific sensitive reference formats.
- Discovery results are documented and reviewed by the data owner before masking proceeds.

5.2.2 Masking Execution

- Apply regex-based or NLP-based redaction to identified sensitive patterns. The redaction pattern must replace the entire sensitive value, not merely flag it.
- For log files: masking is applied at the log writer level (a custom log formatter or log pipeline filter), not post-hoc. Log files written to shared storage must never contain unmasked sensitive data.
- For API payloads: sensitive fields are masked at the API gateway or application layer before responses are logged or forwarded to third-party services. API logging configurations are reviewed as part of the quarterly security audit.
- For documents and email archives: masked copies are produced and stored. Original documents are retained in a restricted access repository unless the document retention policy requires deletion.
- Masking execution and results are logged per the requirements in Section 5.1.3.

5.2.3 Validation

- Sample at least 10% of masked files or 100 files (whichever is greater) to confirm sensitive patterns have been removed.
- Run automated pattern detection across the full masked corpus. Any residual sensitive pattern match is a masking failure.
- Confirm that file structure and non-sensitive content is unaffected by the masking operation.

5.3 Masking in Specific Contexts

5.3.1 Environment Refresh (Production to Non-Production)

- All environment refreshes from production data sources must be routed through an automated masking pipeline before the data is loaded into the target environment.
- The masking pipeline runs in an isolated environment with no production network access after the initial data extract. The pipeline has no ability to write back to the production database.
- A refresh operation that fails masking validation is rejected. The target environment retains its previous dataset until a validated refresh is available.
- Environment refresh masking pipelines are tested at least quarterly by running a validation suite against a sample production extract in a sandboxed environment.

5.3.2 Third-Party Data Sharing

- Data shared with third parties must be masked to the level specified in the technique selection matrix for the Third-party Share column (Section 4.2), regardless of the third party's contractual access rights to the original data.
- A Data Masking Request Form (Appendix A) must be completed and approved by the DPO before any data is prepared for third-party sharing.
- The third party must be notified that the shared dataset is masked and must not attempt to reverse, reconstruct, or combine it with other datasets to re-identify individuals. This obligation is included in the data sharing agreement.
- All data exports for third-party sharing are logged with the recipient identity, dataset description, masking technique applied, and approval reference.

5.3.3 Analytics and Reporting Pipelines

- Analytics pipelines that consume production data must apply masking at the ingestion stage, before the data is written to any analytical store.
- Aggregated outputs that cannot be disaggregated to individual-level data are exempt from masking requirements, provided the aggregation level is sufficient to prevent re-identification (minimum group size of 5 for most contexts; higher thresholds may apply under specific regulatory frameworks).
- Where generalisation is used for analytics, the generalisation intervals must be reviewed by a privacy engineer before the dataset is released.

5.3.4 Support and Helpdesk Systems

- Customer data displayed in support interfaces is masked at the API layer before rendering. Support staff see the masked view by default.
- Unmasking of a specific field for a specific customer interaction requires a logged justification and is time-limited (maximum 1-hour unmasked session per ticket). All unmasked data access events are logged.
- Support ticket content is scanned for sensitive data patterns before the ticket is stored. Any sensitive data identified in ticket content is automatically redacted and the submitting agent is notified.

5.3.5 Local Developer Workstations

- Developer workstations are not trusted environments for the purposes of this procedure. Production data, partially masked data, and production-derived masked datasets must not be copied to or processed on local developer machines under any circumstances.
- Developers requiring realistic data for local development must use centrally generated synthetic datasets. Synthetic datasets contain no production-derived records and require no masking controls beyond format validation.
- Masking pipeline development and testing on local machines must use synthetic source data only. Pipeline configuration referencing any production resource: KMS endpoints, Vault paths, or database connection strings - is a non-conformance under Section 10.2.

5.4 Key and Token Management

- Tokenisation vaults and FPE encryption keys are stored in a hardware security module (HSM) or an HSM-backed cloud key management service (AWS KMS, Azure Key Vault, GCP Cloud KMS, HashiCorp Vault with auto-unseal using HSM). Software-only key storage for production masking keys is a non-conformance.
- Masking keys are classified as Restricted. Key material is never logged, stored in environment variables, or committed to version control.
- Key rotation schedule: FPE keys and tokenisation vault master keys are rotated at a minimum annually or following any suspected compromise. Rotation does not require re-masking of existing datasets if the key management system supports key versioning.
- Key access is role-based: masking pipeline service accounts have encrypt/decrypt permissions only. No human user has direct access to production masking key material except under a documented break-glass procedure with audit logging.
- Token vault databases are separate from application databases. They are not accessible from non-production environments.
- Salt values used for deterministic masking or keyed hashing must be fetched at runtime from the approved KMS or Vault instance via a short-lived service account token scoped read-only to the specific secret path. Salts must not persist in any storage after the masking job terminates. Every retrieval must produce an audit log entry in the KMS.

6. KEY PERFORMANCE INDICATORS

All metrics are sourced from masking pipeline logs, data discovery tooling output, audit records, and the incident management system. Manual collection of any listed metric indicates a tooling gap requiring remediation.

KPI	What it measures	Target	How collected	Breach response
Masking Coverage Rate	Percentage of data assets (tables, object stores, log streams) containing classified sensitive data that have an active, validated masking control in place	100%	Data discovery scan output vs masking control inventory	Coverage below 100% means unmasked sensitive data exists in scope. Each gap is a non-conformance requiring remediation within 5 business days.
Environment Refresh Masking Pass Rate	Percentage of non-production environment refresh operations that complete masking validation without failure before the environment is made available	100%	Masking pipeline execution logs; validation report pass/fail per refresh	Any refresh where masking validation failed but the environment was released anyway is an immediate non-conformance and potential breach.
Automated Pattern Detection False Negative Rate	Percentage of synthetic test records containing known sensitive patterns that are not detected by the automated pattern detection suite	0%	Monthly red-team test: inject known sensitive patterns into a sample masked dataset and run the detection suite	A non-zero false negative rate means the detection suite has coverage gaps. The suite must be updated and re-run within 48 hours of any gap being identified.
PII Re-identification Risk Score	Statistical measure of the re-identification risk of released pseudonymised or generalised datasets (typically k-anonymity value, l-diversity, or t-closeness)	k >= 5 for all quasi-identifier combinations in released analytical datasets; higher thresholds for sensitive attributes	Privacy engineering tooling (ARX, Amnesia, sdcMicro) applied to each released analytical dataset	Any dataset with k < 5 is recalled from the analytical environment and re-generalised before re-release.
Masking Pipeline Execution Time	Elapsed time from masking pipeline trigger to validated dataset availability for the largest	Within defined SLA per environment tier (defined	Pipeline execution timestamps	SLA breach indicates performance degradation in the masking pipeline. Investigate and

	standard environment refresh	in the environment refresh runbook)		remediate within one sprint.
Third-Party Data Share Compliance Rate	Percentage of external data sharing events that have a completed, approved Data Masking Request Form recorded before data release	100%	Data sharing log vs Data Masking Request Form register	Any share event without an approved form is an immediate non-conformance and must be investigated as a potential unauthorised disclosure.
Masking Incident Rate	Number of confirmed masking failures (unmasked sensitive data reaching a non-production environment, third party, or unauthorised user) per calendar quarter	Zero	Incident management system, filtered to masking-related incidents	Any non-zero value triggers a full root cause analysis and an assessment of whether a data breach notification obligation has been triggered.
Key Rotation Compliance Rate	Percentage of masking keys and tokenisation vault master keys that have been rotated within their defined rotation SLA	100%	Key management system rotation timestamp vs defined SLA	Any key outside its rotation SLA is escalated to the CISO and the Information Security team within 24 hours.
Unmasked Data Access Events in Non-Production	Number of logged events where a support or engineering user accessed unmasked data in a non-production environment outside of an approved unmasking session	Zero	Access logs from non-production environments; unmasking session logs	Any non-zero value is investigated as a potential access control failure. The session log is reviewed and the user's access is suspended pending investigation.
Audit Finding Closure Rate	Percentage of masking audit findings closed within their assigned SLA (Critical: 48h; Major NC: 5 business days; Minor NC / OFI: 30 days)	100% within SLA	Audit finding register	Findings outside SLA are escalated to the DPO and reported at the monthly security review.

7. TESTING AND VALIDATION

7.1 Pre-Release Validation Gates

The following validation gates must be passed before any masked dataset is made available in a non-production environment or released to a third party. Gate completion is recorded in the masking execution log.

Gate	Requirement	Evidence Required	Owner
G1: Sensitive Pattern Scan	Automated pattern detection suite returns zero matches against the masked dataset. Patterns must include all data types in Section 3.2.	Scan report with zero findings	Masking pipeline (automated)
G2: Referential Integrity Check	All primary key to foreign key relationships resolve correctly in the masked dataset. No orphaned records.	Integrity check report	DBA / Data Engineer
G3: Format Validation	Masked values conform to the expected format for each column (email format, phone format, date format, numeric range).	Validation report per column	Masking pipeline (automated)
G4: Application Smoke Test	A defined set of smoke tests against the masked dataset pass. Smoke tests must cover all critical data access paths.	Smoke test execution report with pass/fail per test	QA Lead
G5: Row Count Reconciliation	Row counts in the masked dataset match row counts in the source extract. No records lost during masking.	Row count comparison report	Masking pipeline (automated)
G6: Access Control Verification	Access controls on the masked environment are configured per the environment management requirements in QMS-DEV-PRO-006.	Environment access control configuration record	DevOps / Platform Engineer
G7: Approval (Third-party share only)	A completed and approved Data Masking Request Form is on record before the dataset is released to the third party.	Approved form reference	DPO

7.2 Periodic Validation

- Monthly: automated pattern detection suite is run against all active non-production environments containing masked production data. Results are logged and any findings escalate to the masking failure incident procedure in Section 8.
- Quarterly: a red-team test injects synthetic records containing known sensitive patterns into a masked dataset and confirms the detection suite identifies them all. A false negative rate above zero triggers an update to the detection suite within 48 hours.

- Quarterly: a re-identification risk assessment is performed on all analytical datasets released in the previous quarter. Any dataset with k-anonymity below the defined threshold is recalled.
- Annually: a penetration test targeting the masking controls is conducted by the internal security team or an external specialist. The scope includes attempting to reverse tokenisation using intercepted API traffic, attempting to infer original values from masked outputs, and testing the token vault's access controls.

7.3 Performance Testing

- Masking pipeline performance is benchmarked quarterly against the largest standard production extract.
- The benchmark measures: total execution time, peak memory consumption, records processed per second, and CPU utilisation on the masking pipeline host.
- Any degradation of more than 20% from the previous benchmark without a corresponding increase in data volume triggers a pipeline performance review.

8. AUDIT PROGRAMME

8.1 Audit Schedule

Audit Type	Scope	Frequency	Owner	Output
Masking Control Audit	Verify that masking controls are in place for all data assets containing classified sensitive data; confirm technique selection matches the matrix in Section 4.2	Quarterly	Information Security team	Audit report; gaps logged as non-conformances; masking control inventory updated
Masking Pipeline Audit	Review masking pipeline configuration, access controls, version history, and execution logs; confirm pipeline cannot write to production	Quarterly	DevOps / Platform Engineering + Information Security	Pipeline audit report; non-conformances for any pipeline with production write access or unreviewed configuration changes
Key Management Audit	Verify key rotation compliance; confirm keys are stored in approved HSM or KMS; review key access logs for anomalies	Quarterly	Information Security team + CISO	Key management audit report; non-conformances for any key outside rotation SLA

Third-Party Sharing Audit	Verify that all data sharing events in the period have corresponding approved Data Masking Request Forms; review third-party agreements for masking obligations	Semi-annually	DPO + Information Security	Third-party audit report; any sharing event without an approved form is escalated as a potential unauthorised disclosure
Re-identification Risk Assessment	Statistical assessment of all analytical datasets released in the period for re-identification risk; k-anonymity verification	Quarterly	Privacy Engineer or DPO-designated reviewer	Risk assessment report; any dataset below threshold is recalled and re-generalised
Comprehensive Process Audit	Full review of this procedure, its implementation, and compliance with referenced regulatory standards; includes interviews with implementing teams and sampling of masking execution logs	Annually	DPO + Information Security + external auditor (for ISO 27001 / PCI DSS cycles)	Comprehensive audit report; findings with severity; corrective action plan; procedure amendment where required

8.2 Audit Finding Classification and SLAs

Severity	Definition	Closure SLA	Escalation
Critical	Active masking failure: unmasked sensitive data confirmed in a non-production environment, third-party system, or accessible to an unauthorised user. Constitutes a potential data breach.	Immediate containment; full closure within 48 hours	CISO, DPO, Legal; breach notification assessment within 24 hours
Major Non-Conformance	Masking control missing or materially deficient for a Restricted or Regulated data type; key management SLA breached; masking pipeline with production write access; unsupported masking technique in use	5 business days	CISO and DPO; reported at next monthly security review
Minor Non-Conformance	Masking control present but not fully conforming to the procedure (e.g., technique is acceptable but not the preferred choice); documentation gap; minor process deviation	30 days	Information Security team lead; reported at quarterly audit review

Opportunity for Improvement (OFI)	Enhancement to masking controls, tooling, or process that would improve efficiency, coverage, or compliance posture, but where current controls are adequate	Next planning cycle	Engineering Lead or DPO; logged in improvement backlog

9. MASKING FAILURE INCIDENT RESPONSE

9.1 Incident vs. Breach Classification

Classification	Definition	Immediate Action
Masking Failure (Incident)	Masking control failed or was not applied, but no evidence that unmasked data was accessed by an unauthorised party. The data reached a non-production environment or was prepared for sharing in an unmasked state.	Quarantine the affected dataset immediately. Initiate incident record. Assess whether any access has occurred. If no access is confirmed within 2 hours, treat as a Potential Breach.
Potential Breach	Masking failure confirmed AND it cannot be determined whether unauthorised access occurred. The unmasked dataset was accessible to parties who should not have seen unmasked data.	Quarantine. Notify CISO and DPO within 1 hour. Begin breach notification assessment. Preserve all access logs. Assume breach until evidence confirms otherwise.
Confirmed Breach	Masking failure confirmed AND evidence exists that unmasked sensitive data was accessed, downloaded, or transmitted to an unauthorised party.	Quarantine. Notify CISO, DPO, and Legal within 1 hour. Initiate breach notification procedure per applicable regulation (GDPR: 72-hour supervisory authority notification; PCI DSS: notify card brands immediately). Preserve all evidence.

9.2 Incident Response Steps

Step 1: Detection and Reporting

- Any employee who identifies or suspects a masking failure raises an incident immediately via the incident management system, tagging the incident as DATA-MASKING.
- Automated monitoring alerts (pattern detection, access anomaly detection) generate incidents directly in the incident management system.
- Initial severity classification is assigned by the on-call security engineer within 30 minutes of the incident being raised.

Step 2: Containment

- Revoke access to the affected dataset or environment immediately. If the dataset has been shared externally, notify the recipient to quarantine the data and await instructions.

- Suspend any active masking pipeline that produced the failure until the root cause is identified and remediated.
- Preserve all logs, access records, pipeline execution outputs, and database states as forensic evidence. Do not modify or delete any records pending the investigation.

Step 3: Assessment

- Determine: what data was exposed, in what volume, to which parties, for what duration, and whether any access occurred.
- Assess whether the exposure triggers a regulatory notification obligation. GDPR: notify the supervisory authority within 72 hours of becoming aware of a breach likely to result in risk to individuals. PCI DSS: notify the card brands immediately upon discovery of a compromise of cardholder data.
- Document the assessment in the incident record. If notification is required, coordinate with Legal and the DPO.

Step 4: Remediation

- Identify and fix the root cause of the masking failure before any data is re-exposed to the affected environment or third party.
- Re-run the full masking pipeline with the corrected configuration and re-execute all pre-release validation gates (Section 7.1) before the environment or dataset is reinstated.
- Update the masking rule set, pipeline configuration, or detection suite as required to prevent recurrence.

Step 5: Post-Incident Review

- Conduct a post-incident review within 5 business days of incident closure. The review must cover: root cause, detection time, containment time, whether any regulatory notification was required and sent within SLA, and specific corrective actions.
- Post-incident review output is recorded in the incident record and the findings are incorporated into the next quarterly masking control audit.
- If the masking failure reveals a systemic gap in the masking procedure, this procedure is updated via the amendment process defined in Section 1.

10. CONTINUOUS IMPROVEMENT

10.1 Improvement Cadences

Cadence	Trigger / Frequency	Scope	Required Output
Weekly	Weekly	Masking coverage rate; environment refresh pass rate; active incident status	Status report; any metric outside target escalated with named owner and remediation timeline

Monthly	Monthly	Full KPI set in Section 6; audit finding closure status; key rotation compliance; third-party sharing log review	KPI report; finding closure status; escalation of any metric outside target
Quarterly	Quarterly	Masking audit results; red-team test results; re-identification risk assessments; pipeline performance benchmarks; technique selection matrix review	Audit reports; updated masking control inventory; procedure amendments if required
Post-incident	After any masking failure incident	Root cause analysis; detection and containment time vs SLA; regulatory notification assessment; corrective action plan	Post-incident review record; procedure update where systemic gap identified
Annually	Annually	Full procedure review; regulatory landscape review; technique adequacy review (deprecated algorithms, new regulatory requirements); training effectiveness	Updated procedure version; amendment record; re-approval cycle

10.2 Non-Conformance Handling

The following conditions are non-conformances under this procedure and require a documented corrective action within the SLA defined in Section 8.2:

- Unmasked sensitive data confirmed in any non-production environment, third-party system, or support interface without an approved unmasking session record
- An environment refresh where masking validation gates were not fully passed before the environment was made available
- A data sharing event to a third party without a completed, approved Data Masking Request Form on record
- A masking technique applied to a data type that is not the approved technique per the matrix in Section 4.2, without a completed exception form
- A tokenisation vault or FPE key stored outside an approved HSM or cloud KMS
- A masking key outside its rotation SLA without a documented, approved extension
- Use of a prohibited algorithm or technique listed in Section 4.3
- A masking pipeline with write access to any production database
- A biometric data export to any non-production environment or third party under any circumstances

APPENDIX A: DATA MASKING REQUEST FORM

Form Reference: QMS-SEC-FORM-001

Complete this form for all data masking operations. Submit to the Information Security team for review. For third-party sharing, DPO approval is mandatory before data release.

Field	Detail
Request Reference	
Date of Request	
Requestor Name and Role	
Data Owner Name and Role	
Purpose of Masking Operation	
Target Environment or Recipient	Dev / QA / Staging / Analytics / Third Party: _____
Source System / Database	
Data Classification Level	Public / Internal / Confidential / Restricted / Regulated
Data Types to be Masked	List all data types. Reference Section 3.2.
Masking Technique(s) Selected	Reference Section 4.2. Note any deviations and attach exception form (Appendix B).
Estimated Record Volume	
Is third-party sharing involved?	Yes / No. If Yes, recipient name and DPA reference:
Regulatory obligations identified	GDPR / PCI DSS / HIPAA / National DPA / None. Specify:
Data owner approval	Name: _____ Signature: _____ Date: _____
DPO approval (if third-party or Regulated data)	Name: _____ Signature: _____ Date: _____
Information Security approval	Name: _____ Signature: _____ Date: _____
Masking execution completed by	Name: _____ Date: _____
Validation gates passed (Section 7.1)	G1: Y/N G2: Y/N G3: Y/N G4: Y/N G5: Y/N G6: Y/N G7 (if applicable): Y/N
Execution log reference	
Notes / Exceptions	

APPENDIX B: MASKING TECHNIQUE EXCEPTION REQUEST

Form Reference: QMS-SEC-FORM-002

Complete this form when the approved technique for a data type and environment combination (Section 4.2) cannot be applied and an alternative is proposed. Requires Information Security and DPO approval before the exception is implemented.

Field	Detail
Exception Reference	
Date of Request	
Requestor Name and Role	
Data Type(s) Affected	
Target Environment	
Approved Technique per Matrix	Reference Section 4.2
Proposed Alternative Technique	
Justification for Exception	Why cannot the approved technique be applied? Provide technical and/or business justification.
Compensating Controls	What additional controls will be applied to compensate for the deviation?
Risk Assessment	Describe the residual risk of the proposed alternative compared to the approved technique.
Duration of Exception	Date range or condition under which the exception applies
Review Date	Exception must be reviewed on or before this date
Requestor approval	Name: _____ Signature: _____ Date: _____
Information Security approval	Name: _____ Signature: _____ Date: _____
DPO approval	Name: _____ Signature: _____ Date: _____

APPENDIX C: PRE-MASKING EXECUTION CHECKLIST

Form Reference: QMS-SEC-FORM-003

Complete before executing any masking operation. Attach to the Data Masking Request Form (Appendix A) as supporting evidence.

#	Check	Status
1	Data Masking Request Form (QMS-SEC-FORM-001) is completed and all required approvals are obtained	Y / N / N/A
2	Source data extract is isolated from the production database. No masking tooling has write access to production.	Y / N
3	Data discovery has been run and all sensitive columns are identified and classified	Y / N
4	Masking rules are defined for all sensitive columns and match the approved technique per Section 4.2 (or an approved exception per Appendix B is on file)	Y / N
5	Referential integrity dependencies across tables involving masked columns have been mapped	Y / N
6	Substitution reference datasets confirmed to contain no real data	Y / N / N/A
7	Tokenisation vault or FPE key is available and access is confirmed (do not record key material here)	Y / N / N/A
8	Masking pipeline version is recorded and the pipeline has been reviewed per the Code Review Procedure	Y / N
9	Target environment access controls have been confirmed as correctly configured before data is loaded	Y / N
10	Masking execution log is configured and will capture all required fields (timestamp, operator, tables, techniques, row counts)	Y / N
11	Rollback plan is defined: if masking fails mid-execution, the partially masked dataset will be discarded and the target environment will not receive any data	Y / N
12	Validation gates (Section 7.1) are scheduled to run immediately on completion of masking execution	Y / N

Completed by: _____ Date: _____ Signature: _____

APPENDIX D: MASKED DATA VALIDATION CHECKLIST

Form Reference: QMS-SEC-FORM-004

Complete after masking execution and before the masked dataset is made available. All gates must pass before the dataset is released. Attach to the Data Masking Request Form.

#	Validation Check	Evidence / Tool	Pass / Fail
G1	Automated pattern detection suite returns zero findings for all sensitive data types across the full masked dataset	Scan report reference	P / F
G2	All primary key to foreign key relationships resolve without orphaned records	Integrity check report	P / F
G3	Masked field values conform to expected format per column type (email, phone, date, numeric)	Validation report	P / F
G4	Application smoke test suite passes at the required threshold against the masked dataset	Test execution report	P / F
G5	Row count in the masked dataset matches the row count in the source extract (no record loss)	Row count comparison	P / F
G6	Access controls on the target environment are confirmed as correctly configured	Configuration record	P / F
G7 (Third-party only)	Approved Data Masking Request Form is on file before data is released	Form reference	P / F / N/A
Add'l	For analytical datasets: re-identification risk assessment completed and k-anonymity >= 5 for all quasi-identifier combinations	Risk assessment report	P / F / N/A

Overall result: PASS / FAIL

If any gate fails: do not release the dataset. Record the failure, identify the root cause, correct the masking pipeline, and re-execute from the beginning.

Validated by: _____ Role: _____Date: _____ Signature: _____

APPENDIX E: MASKING FAILURE INCIDENT REPORT TEMPLATE

Form Reference: QMS-SEC-FORM-005

Raise immediately on identification of a masking failure. Submit to the Information Security team and the DPO within 1 hour of identification.

Field	Detail
Incident Reference	
Date and Time of Detection	
Detected by (Name and Role)	
Incident Classification	Masking Failure / Potential Breach / Confirmed Breach (circle one)
Description of the failure	What masking control failed and how was the failure detected?
Data types exposed	List all data types that may have been exposed in unmasked form
Data classification level	
Volume of records potentially affected	
Systems / environments affected	
Duration of exposure (if known)	From: _____ To: _____
Parties who may have accessed unmasked data	
Containment actions taken	Describe immediate containment steps and timestamp of each action
Is regulatory notification required?	GDPR 72h / PCI DSS immediate / National DPA: _____ / Not triggered (justify)
Notification deadline (if applicable)	
Root cause (preliminary)	Complete within 24 hours of containment
Remediation steps taken	
Validation that masking failure is resolved	Reference validation gate results (Section 7.1)
Post-incident review scheduled date	
Reported by	Name: _____ Signature: _____ Date: _____
CISO review	Name: _____ Signature: _____ Date: _____
DPO review	Name: _____ Signature: _____ Date: _____

APPENDIX F: GLOSSARY OF TERMS

Term	Definition
AES-FF1 / AES-FF3-1	NIST-approved format-preserving encryption modes defined in NIST SP 800-38G. AES-FF1 is preferred for new implementations.
Anonymisation	The irreversible transformation of personal data such that the data subject is no longer identifiable directly or indirectly. Anonymised data is not personal data under GDPR.
Argon2id	A memory-hard password hashing algorithm recommended by OWASP and NIST for credential storage. Resistant to GPU-based brute force attacks.
bcrypt	A widely used password hashing function incorporating a work factor to slow brute force attacks. Acceptable for credential storage where Argon2id is not available.
Data Masking	The process of replacing sensitive data with structurally similar but non-sensitive values, reducing exposure while preserving the usability of the data for its intended purpose.
FPE (Format-Preserving Encryption)	An encryption technique that produces ciphertext in the same format as the plaintext input. See AES-FF1 / AES-FF3-1.
Generalisation	The replacement of a specific value with a less specific but accurate range or category, used to reduce re-identification risk in analytical datasets.
HSM (Hardware Security Module)	A physical computing device that safeguards and manages cryptographic keys and provides hardware-accelerated cryptographic operations.
k-Anonymity	A property of a dataset where each combination of quasi-identifying attributes appears in at least k records, limiting re-identification risk.
KMS (Key Management Service)	A cloud-provider-managed service for creating, storing, and controlling access to cryptographic keys (AWS KMS, Azure Key Vault, GCP Cloud KMS).
PAN (Primary Account Number)	The payment card number printed on a card or encoded on its magnetic stripe, as defined by ISO/IEC 7812.
PII (Personally Identifiable Information)	Any information that can be used to identify a specific individual, directly or in combination with other data.
Pseudonymisation	The replacement of direct identifiers with pseudonyms, with the mapping stored separately. Data remains personal data under GDPR.
Re-identification	The process of combining anonymised or pseudonymised data with other datasets to recover the identity of individuals. A key risk in the release of analytical datasets.
SAD (Sensitive Authentication Data)	Payment card data that must not be stored after authorisation under PCI DSS, including CVV/CVC, full magnetic stripe data, and PIN blocks.
Tokenisation	The replacement of a sensitive value with a non-sensitive token, with the original value stored in a secure token vault. The token has no mathematical relationship to the original value.