

A short primer on data classification

Deciding how sensitive your data is — and therefore how hard to protect it.

READING TIME ~7 minutes · BUILDS ON DOCUMENT 2 (ASSETS) · FEEDS YOUR IMPACT SCORES

Once you've listed your assets, the next question is: *how much does each one matter?* Not all data is equal — a public blog post and a database of patient records need very different protection.

Data classification is how you make that difference explicit, and it turns out to be the bridge between your asset list and the *impact* scores you'll give risks in the workshop.

What data classification is

Data classification is the process of sorting information into categories according to how sensitive it is and how much protection it needs. It's not optional busywork: ISO/IEC 27001:2022 makes it a required control (**Annex A 5.12, Classification of information**), which says information should be classified by its **confidentiality, integrity and availability** needs, together with legal requirements, its value, criticality and sensitivity.^[1]

IN ONE SENTENCE

Classification answers “**how bad would it be if this data leaked, was tampered with, or became unavailable?**” — and gives that answer a name everyone can act on.

Classify on three questions: C-I-A

You classify a piece of information by asking what the impact would be across the three security objectives — the same C-I-A lens you met in the assets guide:

- **Confidentiality** — how damaging if the wrong people saw it?
- **Integrity** — how damaging if it were wrongly altered or corrupted?
- **Availability** — how damaging if it were unavailable when needed?

THE HIGH-WATER MARK

NIST's FIPS 199 rates each objective as **Low, Moderate or High** impact — and the **highest** of the three sets the overall level.^[3] If a dataset is Low for availability but High for confidentiality, it's classified **High**. Protect to the worst case.

A simple four-level scheme

ISO doesn't prescribe the exact labels — each organisation defines its own. A common, practical scheme has four levels. (Fewer is usually better than more: the guidance warns against over- and under-classifying.^[1])

Level	Use it for	Example (software / health context)	If exposed
Public	Cleared for anyone to see	Marketing site, public API docs, open-source code	No harm
Internal	Default for everyday work	Internal wikis, tickets, non-sensitive configuration	Minor harm
Confidential	Sensitive; limited access	Source code, business plans, customer contact details	Serious harm
Restricted	Most sensitive; strict need-to-know	Patient/health records, secrets & keys, special-category data	Severe harm / legal exposure

Label it, then handle it

A classification is only useful if it's visible and acted upon. ISO/IEC 27001 pairs classification with **Annex A 5.13, Labelling of information**: once data has a level, label it and apply the handling rules that match.^[1] Higher levels earn stricter handling:

- **Storage** — encryption at rest for Confidential and above
- **Transmission** — encryption in transit (TLS); no sensitive data over open channels
- **Access** — need-to-know, least privilege, role-based access
- **Sharing** — approval or agreements before Confidential data leaves the organisation
- **Retention & disposal** — keep only as long as needed, then delete securely

Personal data is a special case

If information identifies a living person, data-protection law raises the stakes. Under the EU GDPR, ordinary **personal data** already needs care — and **special categories** (health, genetic, biometric, and more) get extra legal protection.^[5] For a health product, patient data is special-category by definition, so it should default to your highest classification. This is exactly why the Assets tab has a dedicated *Personal data* column: a “Yes” there is a strong signal to classify high.

Classify in four steps

- 1 **Identify the information** this asset holds — what is it, and who does it describe?
- 2 **Assess the impact** on Confidentiality, Integrity and Availability if it were compromised (Low / Moderate / High).
- 3 **Assign the level** — take the highest of the three (the high-water mark).
- 4 **Label it and apply the handling rules** that match that level.

WORKED EXAMPLE

Patient records. Confidentiality → **High** (health data, GDPR special category). Integrity → **High** (clinical decisions depend on it). Availability → **Moderate**. High-water mark = High → classify **Restricted**, encrypt at rest and in transit, strict need-to-know access.

How this connects to the workshop

Classification is essentially the **Impact** half of Likelihood × Impact. When you score a risk to a data asset, its classification tells you how high the impact can go — a leak of *Restricted* data is a 5, a leak of *Public* data is a 1. Capture the signal on the Assets tab via *Type of information* and *Personal data*, and let the higher-classified assets pull your attention first in the register.

References & further reading

1. ISO/IEC 27001:2022, Annex A controls 5.12 *Classification of information* and 5.13 *Labelling of information*. ISO. [iso.org/standard/27001](https://www.iso.org/standard/27001)
2. ISO/IEC 27002:2022 — *Information security controls* (implementation guidance for 5.12 / 5.13). ISO. [iso.org/standard/75652.html](https://www.iso.org/standard/75652.html)
3. NIST FIPS 199 — *Standards for Security Categorization of Federal Information and Information Systems* (Low/Moderate/High impact). NIST. csrc.nist.gov/pubs/fips/199/final
4. NIST SP 800-60 Vol. 1 Rev. 1 — *Guide for Mapping Types of Information and Information Systems to Security Categories*. NIST. csrc.nist.gov/pubs/sp/800/60/v1/r1/final
5. Regulation (EU) 2016/679 (GDPR), Article 9 — *Processing of special categories of personal data*. gdpr-info.eu/art-9-gdpr/
6. ISMS.online — *ISO 27001:2022 Annex A Control 5.12 explained*. isms.online/iso-27001/annex-a-2022/5-12-classification-of-information-2022/

Standards summaries are paraphrased for teaching; consult the official ISO/IEC and EUR-Lex texts for authoritative wording. NIST publications are US Government works in the public domain.