23 Data protection along the entire data flow

AUTHOR
R. Heller

R. Heller · Date: 2025-10-19 · Last changed: 2026-11-09

At a glance

Guiding question: Who receives which data, for what purpose, and for how long?
For: clinical researchers, data stewards, teachers who want to show real examples, and reviewers; the core reading needs no programming. · **Prerequisites:** the gait measurement from video in Chapter 11; how data arise in Chapter 3; the regulatory note in Chapter 22 helps but is not required.
In the course: W7 · Regulation, data protection and incident (data flow and DPIA screening); also used in W1, W2, W4 and W8 · **Data:** synthetic (data-flow register generated in the chapter); legal status checked on 8 November 2026 · **Core reading time:** about 11 minutes

23.1 Case and decision

The rehabilitation clinic of Chapter 11 films patients walking after surgery for a tibial plateau fracture. The videos support treatment decisions. A research team wants to compare video-based knee angles with the marker-based reference method, and a lecturer wants one clip for a seminar. An external service provider offers cloud pose estimation that turns video into body keypoints (hip, knee, ankle and other landmarks) within minutes. By Friday the study lead has to decide what leaves the clinic: the raw videos, the keypoints, or only the angles. She must also say who holds the list linking study numbers to patients, and when each copy is deleted. She decides that keypoints are extracted on site, that the provider receives no video, and that the seminar uses a synthetic clip. The study starts three weeks later than planned, but no copy exists whose purpose and deletion date are unknown.

23.2 Core reading

Core reading · required for W7 · about 11 minutes · ends before ‘Decision exercise’

Follow the copies, not the project. Data protection questions become answerable when every copy of the data is drawn as a step with a purpose, a holder and an end. Figure 23.1 traces a gait video from the camera to a knee angle in a research table. It marks where people are directly identifiable, where data are pseudonymised, where the boundary of the controlling institution is crossed and where each copy is deleted. All statements on law refer to the GDPR (Regulation (EU) 2016/679) and German law as of 8 November 2026 (European Parliament and Council of the European Union 2016).

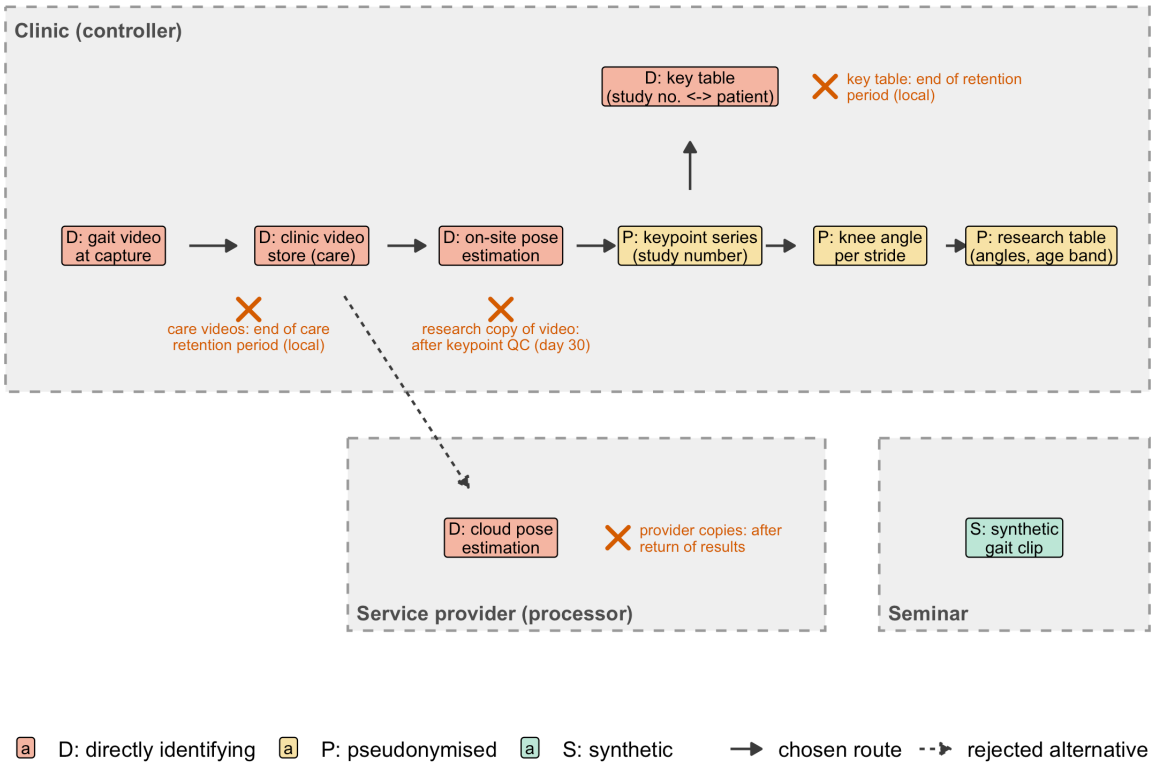


Figure 23.1: Data flow from gait video to knee angle in the fictional case. Boxes are copies or processing steps; the prefix gives the identifiability: D = directly identifying, P = pseudonymised (still personal data), S = synthetic. Shaded areas are organisational boundaries. Crosses mark deletion points; days are assumptions to be set locally. Fictional flow; no data.

Two legal questions, asked separately. A gait video recorded for treatment shows a person's body, face and condition. It is personal data and, because it reveals information about health, data concerning health. Each processing step therefore needs two answers that must not be merged. First, a lawful basis under Art. 6(1) GDPR, for example the performance of a treatment contract or a task in the public interest [Law]. Second, an exception from the prohibition on processing special categories under Art. 9(2), for example medical diagnosis and treatment under a duty of professional secrecy (Art. 9(2)(h) with 9(3)) or research based on Union or Member State law with suitable safeguards (Art. 9(2)(j)) [Law]. A step with an Art. 6 basis but no Art. 9 exception is unlawful, and the reverse as well. As in Chapter 22, the labels [Law], [Guidance], [Interpretation] and [Local] mark the kind of statement.

Care, research and teaching are different purposes. Table 23.1 compares them. Research has its own routes: consent (Art. 6(1)(a) with 9(2)(a)) or a statutory research provision under Art. 9(2)(j). In Germany, the Federal Data Protection Act (Bundesdatenschutzgesetz, BDSG), § 27, permits processing special categories for research without consent if it is necessary and the controller's interest substantially outweighs the person's interest in exclusion, with specific safeguards [Federal Republic of Germany (2018); Law]. The BDSG applies to federal public bodies and to private bodies; to public bodies of the federal states it applies only in narrow cases where state law does not regulate data protection (§ 1(1)) [Law]. Whether a hospital falls under § 27 BDSG, a state hospital or data protection act, or the Health Data Use Act (Gesundheitsdatennutzungsgesetz,

GDNG), § 6, which lets health institutions further process their own care data for research if the data are pseudonymised and personal data are generally not passed to third parties [Federal Republic of Germany (2024); Law], is a [Local] question for the data protection officer. **The key assumption: calling an activity "research" exempts nothing by itself; a specific legal provision with its conditions has to fit the specific step.** Teaching is not research, so no research provision covers the seminar clip [Interpretation].

Table 23.1: Care, research and teaching use of the same gait video. Fictional case; legal status 8 November 2026; bracketed labels mark interpretation and local questions.

	Care	Research	Teaching
Purpose	Treatment decision for this patient	Agreement of two angle measurements	Showing students what a gait pattern looks like
Art. 6 GDPR	Treatment contract or legal obligation (Art. 6(1)(b), (c))	Consent or public task (Art. 6(1)(a), (e))	Consent (Art. 6(1)(a))
Art. 9 GDPR	Art. 9(2)(h) with professional secrecy (9(3))	Art. 9(2)(a) consent, or 9(2)(j) with national law	Art. 9(2)(a) explicit consent
National law	Professional and hospital law [Local]	§ 27 BDSG, state law or § 6 GDNG – whichever applies [Local]	None that replaces consent [Interpretation]
Who sees raw video	Treating team	Nobody outside the on-site extraction step	Nobody: synthetic clip
Deletion	End of care retention period [Local]	Research video copy after keypoint QC; key table at end of retention period	Not applicable for synthetic data

Pseudonymisation is not anonymisation. The GDPR defines pseudonymisation as processing such that data can no longer be attributed to a specific person without additional information that is kept separately and protected (Art. 4(5)) [Law]. The additional information here is the key table. It needs a named holder (for example a data custodian outside the research team), a written rule for when re-identification is allowed (withdrawal of consent, an incidental finding) and a deletion date. Pseudonymised data remain personal data for the controller. The Court of Justice held in 2025 that pseudonymised data need not be personal data "in all cases and for every person": for a recipient who cannot lift the pseudonymisation and cannot identify the person by other means, including by cross-checking with other factors, they may not be; but the controller's duty to inform people about recipients is assessed at collection, from the controller's point of view (Court of Justice of the European Union 2025, paras 77, 86, 111–112). The judgment concerned the parallel regulation for EU institutions; its transfer to the GDPR is plausible but remains [Interpretation]. For gait it matters a great deal: movement itself can identify. In one study, 95% of 511 people were recognised from less than five minutes of virtual-reality motion tracking data [Miller et al. (2020); *published*]. Keypoint series should therefore be treated as personal data at every recipient unless the opposite has been shown.

Roles and borders. The clinic, as the body that decides on purposes and means, is the controller (Art. 4(7)) [Law]. A cloud service that estimates poses on its instructions is a processor; this requires a contract that, among other things, binds it to documented instructions and to deletion or return of data at the end (Art. 28(3)) [Law]. If the provider stores data or provides support from outside the European Economic Area, the transfer rules of Chapter V apply: an adequacy decision

(Art. 45) or appropriate safeguards such as standard contractual clauses (Art. 46) [Law]. People keep their rights to information, access, rectification, erasure and objection (Art. 13–21); for research, restrictions exist only where Union or Member State law provides them (Art. 89(2); § 27(2) BDSG) [Law].

When is a DPIA needed? A data protection impact assessment is required where processing is likely to result in a high risk, in particular for large-scale processing of special categories (Art. 35(1), 35(3)(b)) [Law]. “Large scale” has no number in the text; the supervisory authorities publish lists of processing that requires a DPIA (Art. 35(4)) [Law; Local]. New technology, health data and a cloud provider make a documented screening the minimum [Interpretation].

Question for the workshop: Your clinic wants to keep the raw research copy of each video “in case a better pose model appears”. Which purpose, legal basis and deletion date would you have to write down to allow this, and would you sign it?

23.3 Decision exercise

Role task. Work in four roles: study lead, data protection officer, clinic IT and a patient representative. Trace the data flow of Figure 23.1 step by step. For each step write down: data, identifiability, purpose, Art. 6 basis, Art. 9 exception, role of the holder, location, transfer mechanism (if outside the EEA), deletion point and residual re-identification risk. The patient representative may veto any step whose purpose cannot be explained in one sentence.

A register that checks itself. The register can be written as a small table in R with explicit rules. This does not replace a legal assessment; it makes omissions visible.

```
# Data-flow register for the fictional gait study (one row per copy or processing step)
register <- data.frame(
  step      = c("capture", "clinic_store", "pose_onsite", "keypoints", "angles",
       "research_table",
               "key_table", "teaching_clip"),
  level     = c("direct", "direct", "direct", "pseudonymised", "pseudonymised",
       "pseudonymised",
               "direct", "synthetic"),
  purpose   = c("care", "care", "research", "research", "research", "research", "research",
       "teaching"),
  art6      = c("6(1)(b)", "6(1)(c)", "6(1)(a)", "6(1)(a)", "6(1)(a)", "6(1)(a)", "6(1)
       (a)", NA),
  art9      = c("9(2)(h)", "9(2)(h)", "9(2)(a)", "9(2)(a)", "9(2)(a)", "9(2)(a)", "9(2)
       (a)", NA),
  holder    = c("clinic", "clinic", "clinic", "clinic", "clinic", "clinic", "data
       custodian", "lecturer"),
  processor = c(FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FALSE),
  art28     = c(NA, NA, NA, NA, NA, NA, NA, NA),
  location  = c("EU", "EU", "EU", "EU", "EU", "EU", "EU", "EU"),
  chapter_v = c(NA, NA, NA, NA, NA, NA, NA, NA),
  deletion  = c("after transfer to store", "care retention period (local)", "after keypoint
       QC, day 30",
               "end of study", "end of study", "retention period (local)", "retention
       period (local)", NA),
  stringsAsFactors = FALSE
)

# Rules: each returns the steps that violate it
```

```
check_register <- function(r) {
  personal <- r$level != "synthetic"
  list(
    missing_art6        = r$step[personal & is.na(r$art6)],
    missing_art9        = r$step[personal & (is.na(r$art9) | r$art9 == "none")],
    processor_no_art28  = r$step[r$processor & (is.na(r$art28) | r$art28 != "contract")],
    third_country_no_v  = r$step[r$location != "EU" & is.na(r$chapter_v)],
    personal_no_delete  = r$step[personal & is.na(r$deletion)],
    teaching_real_data  = r$step[r$purpose == "teaching" & personal & r$art9 != "9(2)(a)"]
  )
}
result <- check_register(register)
result
```

```
#> $missing_art6
#> character(0)
#>
#> $missing_art9
#> character(0)
#>
#> $processor_no_art28
#> character(0)
#>
#> $third_country_no_v
#> character(0)
#>
#> $personal_no_delete
#> character(0)
#>
#> $teaching_real_data
#> character(0)
```

```
# Expected result: no rule is violated
stopifnot(all(lengths(result) == 0))
```

Expected result. A register in which every personal-data step has both an Art. 6 basis and an Art. 9 exception, every copy has a deletion point, the key table has a named holder and the seminar uses synthetic data. Learners also write one sentence on the residual risk: “Keypoint series and angles remain re-identifiable through movement patterns and visit dates; they are shared only within the clinic.”
Alternative. The cloud provider is used under an Art. 28 contract with EU-only processing and deletion after return of results. This is lawful in principle but adds a recipient, a contract to monitor and a possible transfer question if support staff work from outside the EEA.
Residual question. Who checks, a year from now, that the research copy of each video really was deleted on day 30?

The incident clock. In the fictional incident of Chapter 24, the service provider tells the hospital on a Monday morning that a radiology report excerpt appeared in its support log. A processor must inform the controller without undue delay (Art. 33(2)); the controller notifies the supervisory authority without undue delay and, where feasible, within 72 hours of becoming aware, unless the breach is unlikely to result in a risk to people’s rights and freedoms (Art. 33(1)) [Law]. The EDPB considers a controller “aware” when it has a reasonable degree of certainty that a security incident has led to personal data being compromised [European Data Protection Board (2023), para. 31; Guidance].

```
# Fictional times (Europe/Berlin); awareness = reasonable certainty after the first check
provider_message <- as.POSIXct("2026-11-09 07:40", tz = "Europe/Berlin")
aware            <- as.POSIXct("2026-11-09 10:30", tz = "Europe/Berlin")
deadline_33      <- aware + 72 * 3600
# English day and month names regardless of the computer's locale
fmt_time <- function(x) {
  sprintf("%s %s %s %s, %s", c("Mon", "Tue", "Wed", "Thu", "Fri", "Sat", "Sun")
        [as.integer(format(x, "%u"))],
          format(x, "%d"), month.abb[as.integer(format(x, "%m"))], format(x, "%Y"),
        format(x, "%H:%M"))
}
data.frame(event = c("processor informs controller", "controller aware", "72 hours end"),
           time = fmt_time(c(provider_message, aware, deadline_33)))
```

```
#>                          event                   time
#> 1 processor informs controller Mon 09 Nov 2026, 07:40
#> 2             controller aware Mon 09 Nov 2026, 10:30
#> 3                 72 hours end Thu 12 Nov 2026, 10:30
```

```
# The 72 hours run from awareness, not from the end of the analysis: Thursday, 10:30
stopifnot(format(deadline_33, "%u") == "4", deadline_33 - aware == as.difftime(72, units =
        "hours"))
```

The clock is only one part of the decision. Three further questions must be answered by named people, not by code: is a risk to the people affected unlikely (then no notification, but documentation under Art. 33(5)); is the risk likely to be high (then the people must also be informed under Art. 34(1), unless one of the conditions of Art. 34(3) applies, for example data rendered unintelligible by encryption); and if notification comes after 72 hours, what are the reasons for the delay? [Law]

23.4 Counter-check

Change one assumption and predict which rule fails before you run the code. Version 1: the clinic chooses the cloud provider, whose support team works from a third country, and the contract has not yet been signed. Version 2: the lecturer replaces the synthetic clip with a real patient video “because it is more convincing”, without consent.

```
# Version 1: cloud pose estimation instead of on-site extraction
v1 <- register
v1$holder[3]    <- "cloud provider"
v1$processor[3] <- TRUE
v1$location[3]  <- "third country"
# Version 2: real video in the seminar without consent
v2 <- register
v2$level[8] <- "direct"; v2$art6[8] <- "6(1)(e)"; v2$art9[8] <- "none"; v2$deletion[8] <-
        "end of term"
failed <- function(r) names(Filter(length, check_register(r)))
list(version_1 = failed(v1), version_2 = failed(v2))
```

```
#> $version_1
#> [1] "processor_no_art28" "third_country_no_v"
#>
#> $version_2
#> [1] "missing_art9"       "teaching_real_data"
```

```
stopifnot(identical(failed(v1), c("processor_no_art28", "third_country_no_v")),
          identical(failed(v2), c("missing_art9", "teaching_real_data")))
```

Version 1 fails twice: there is no processor contract and no Chapter V mechanism. Both can be repaired, but only before the first video leaves the clinic. Version 2 fails twice: the step has no Art. 9 exception, and a real patient video in teaching needs explicit consent; no research provision covers a seminar. A public-task basis for teaching (Art. 6(1)(e)) would still lack an Art. 9 exception [Interpretation]. If the rule set had only one column “legal basis”, version 2 would have passed. This is the practical reason for keeping Art. 6 and Art. 9 apart.

23.5 Consequences for patients and operations

An open question changes the plan. Whether keypoint series are personal data for the cloud provider cannot be settled without knowing what else the provider holds. The judgment of 2025 makes the answer depend on the recipient’s actual means of identification (Court of Justice of the European Union 2025), and the motion study suggests that those means may be small (Miller et al. 2020). The study team therefore plans as if they were personal data. Table 23.2 shows the effect.

Table 23.2: Deployment plan for the gait study before and after the data-protection review. Fictional plan.

Element	Plan before the review	Plan after the review
Pose estimation	Cloud service, video upload	On-site extraction; provider receives no video
Raw research copy	Kept “for later models”	Deleted after keypoint quality control (day 30, assumption)
Key table	In the study folder	Held by a data custodian; re-identification only for withdrawal or incidental findings
Seminar	Real patient clip	Synthetic gait clip; real clips only with explicit consent and a deletion date
DPIA	“Not needed for a small study”	Documented screening; full DPIA if the supervisory authority’s list or the data volume requires it
Information sheet	Generic	Names recipients, purposes, retention and the contact for withdrawal

For patients, the change means that their face and walk do not leave the clinic and that they can find out who holds a copy. **Ethics** (Chapter 21): patients with rare injuries or visible impairments are easier to recognise; a small group needs more protection, not less. **Regulation** (Chapter 22): if the angle is used for treatment decisions, the software may become a medical device. **Security** (Chapter 24): the key table and the video store are the assets whose confidentiality matters most.

> ⓘ **The incident from the data-protection perspective**
>
> Which personal data were accessible to whom, who assesses the risk, and who decides on notification and communication? The controller decides, advised by the data protection officer; the processor only reports. The assessment starts from facts: which report excerpts, how many people, who could read the support log, and whether the log has been deleted at the provider. A notification under Art. 33 is required

unless a risk is unlikely; it should be made, where feasible, within 72 hours of awareness, and may be completed in phases (Art. 33(4)). Informing the people affected is a separate decision with a higher threshold – a likely high risk – and its own exceptions (Art. 34) [Law]. The technical restart (Chapter 24) has a different trigger, a different owner and a different end point; a good handover names both.

23.6 Look up and transfer

Reviewer question. Does the manuscript state which data left the institution, in which form, to whom and on which legal basis, and how long copies were kept? “Data were pseudonymised” without saying who held the key, and whether recipients could re-identify, is not a description of a safeguard.

Self-test. The supplier of the fracture alert in Chapter 9 asks the emergency department to send “the original DICOM images” of 30 false alarms so that it can check the model. What do you send, and what must be in place first?

Answer

Annotated sources.

- European Parliament and Council of the European Union (2016) – the GDPR, from Art. 4 definitions to Art. 33–35; the binding text for every step in the register.
- Federal Republic of Germany (2018) and Federal Republic of Germany (2024) – two German research provisions with different addressees and conditions; neither applies automatically.
- Court of Justice of the European Union (2025) – the 2025 judgment on pseudonymised data at the recipient and on the controller’s duty to inform.
- Miller et al. (2020) – published evidence that movement data identify people.
- European Data Protection Board (2023) – EDPB guidance on breach notification, including when a controller becomes “aware”.
- European Data Protection Board (2025) – EDPB guidance on pseudonymisation and the “pseudonymisation domain” (version for public consultation).

23.7 Further reading

- Chapter 11 – how the knee angle is computed from keypoints and why raw video may still be needed for quality control.
- Chapter 3 – how small, rare groups in a table make people recognisable.
- Chapter 24 – protecting the key table, the video store and the logs.
- Chapter 26 – secondary use, controlled access and the stepwise application of the European Health Data Space Regulation, whose chapter on secondary use applies from 26 March 2029 (Art. 105) (European Parliament and Council of the European Union 2025).
- Methods in R: The research workflow – where data acquisition and its documentation sit in a project.

- Scientometrics in R: Reproducible pipelines – documenting each processing step so that a register can be kept up to date.
- CTTIR Courses: Research workflow – planning data handling before the first measurement.
- CTTIR Courses: Import, joins, and missingness with dplyr – the joins that a key table makes possible, and therefore why it must be guarded.