

THE TRAKMY GUIDE · DATA & AI

Getting your asset data ready for AI

How to move from a connected fleet to genuinely usable data: the asset register, six quality dimensions, governance, architecture and a 90-day roadmap to turn your fleet into a reliable data source.

85%

OF AI PROJECTS FAIL (ON DATA)

6

QUALITY DIMENSIONS

90 days

ROADMAP

01 The data, not the model

Why 85% of AI projects fail

02 What "AI-ready" data means

5 properties, 3 states of maturity

03 The asset register

The foundation everyone overlooks

04 The 6 quality dimensions

Measure, monitor, correct

05 Telemetry

Time series, position, sensors

06 Governance, GDPR, sovereignty

Who owns it, who accesses it, where it lives

07 The data architecture

From collection to decision

08 The AI use cases

Surplus fleet, anomalies, maintenance

09 The 90-day roadmap

Frame, stabilise, exploit

10 Pitfalls & checklist

Checking that your fleet is ready

● EXECUTIVE SUMMARY

The model is almost never the weak link

STRAIGHT ANSWER

Asset data that is "AI-ready" is not abundant data: it is data aligned to a use case, uniquely identified, governed, fed by controlled pipelines and continuously verified.

85%

Of AI failures = data (Gartner)

> 80%

Of AI projects never reach production
(RAND)

42%

Abandoned AI in 2025 (S&P)

The real problem

Fitting trackers to 300 machines creates a pool of data — but until it is structured, identified and governed it remains "a dot on a map". The problem is almost never the absence of data, but its **state**.

A special case

Fleet data is **temporal** (the moment and the order matter), **spatial** (referenced to zones) and **large and continuous**. Three properties that conventional data management does not cover.

The hidden cost

More than a quarter of organisations lose over \$5m a year to poor data quality (IBM, 2025): decisions taken on a false utilisation rate, surplus fleet retained, alerts ignored, maintenance that predicts nothing.

"Investing in the data before the algorithm inverts the failure ratio."

Sources: Gartner 2025; RAND Corporation; S&P Global 2025; IBM Institute for Business Value 2025. Dated external figures, indicative.

What "AI-ready" data actually is

STRAIGHT ANSWER

AI-ready data is *fit for purpose*: neither the largest volume nor the freshest — it is data a model or a dashboard can use **without manual reprocessing**.

The five properties to check, one by one

- ✓ **Aligned to a use case**: you know which question it answers (surplus fleet, breakdown, route, evidence).
- ✓ **Governed at asset level**: every asset has a unique identity, an owner and access rules.
- ✓ **Fed by pipelines with control points**: validity checked at every stage.
- ✓ **Described by living metadata**: origin, meaning, unit, date — otherwise a value is just an orphan number.
- ✓ **Continuously assured**: quality is monitored permanently, not verified once.

Three states of maturity

Raw data

The frame that comes out of the tracker: technical ID, coordinates, timestamp, sensors. **Accurate but mute.**

Structured data

Reconciled against the register: named asset plus category, presence in a zone, normalised time. **Usable.**

Enriched data

Cross-referenced with other sources (finance, scheduling, maintenance). **Ready for AI** — this is where the value sits.

A SIMPLE TEST

"If I handed this file to an analyst who does not know my fleet, could they use it without asking me anything?" If not, the data is not ready — and a model is even less forgiving.

The asset register

STRAIGHT ANSWER

If you do only one thing before talking about AI: build and maintain a clean register. Almost every project stumbles on the absence of this foundation — the data exists, but it is attached to nothing.

A pivot identifier: neutral and immutable

Every asset has a unique identifier that **never** changes, even if it is reassigned or re-equipped. All other identifiers (serial number, finance code, registration plate) attach to it through a mapping table. The tracker changes? You update the mapping and the history stays intact.

A closed taxonomy

Category, sub-category, model: chosen from a **closed list**, never free text. "Digger / excavator / mini-excavator" describes the same machine while making it impossible for any process to group them.

Attribute	Why it unlocks analysis
Pivot identifier	Reconciles sources, guarantees history
Category (closed list)	Enables aggregation across families
Acquisition value	Calculates ROI, weights alerts
In-service date	Calculates age, the basis of maintenance
Assignment / site	Detects gaps between theory and the field
Status	Distinguishes normal downtime from an anomaly

WARNING SIGN

A register that lives in a shared spreadsheet, updated by hand with no rules, is unreliable by default. A serious register has a **named owner**, entry rules, a history and a single source of truth.

The six quality dimensions

STRAIGHT ANSWER

"Quality" is too vague to act on. It breaks down into six measurable dimensions. One failing dimension is enough to skew a model — AI propagates errors rather than correcting them.

Dimension	What it measures	How to control it
Completeness	Share of expected data actually present	% per asset and per period; suspicion threshold
Accuracy	Gap between the data and reality	Plausibility rules, cross-checks, field sampling
Consistency	Absence of contradictions within and between sources	Cross-checks; stable semantics and units
Timeliness	Delay from production to availability	Match the cadence to the need (≠ as fresh as possible)
Uniqueness	No duplicate asset, event or measurement	Pivot identifier + ID and timestamp deduplication
Traceability	You know where data came from and how it changed	Lineage metadata all along the pipeline

THE TAKEAWAY

Weak dimensions do not offset each other: a model is limited by the worst one. Build the **quality profile** of your fleet and fix first whatever falls below threshold (usually consistency and traceability).

Time series, position and sensors

STRAIGHT ANSWER

Telemetry describes how the asset actually behaves. It is the most delicate data: continuous, irregular and sensitive to field conditions.

- ✓ **Normalise time:** everything against a single reference (UTC stored, local display). A time series with the wrong time is unusable.
- ✓ **Handle the gaps:** offline storage to backfill later; **flag** periods with no signal, and never interpolate silently.
- ✓ **Position ≠ presence in a zone:** keep the nature (GPS, Wi-Fi, BLE, cellular) and the accuracy of every point.

Measurement	Relevant cadence
Position (passive asset)	1 to 4 / day + movement event
Position (real time)	Continuous while moving
Activity / vibration	Sampled (threshold + aggregation)
Shock	Event-driven (threshold)
Temperature / humidity	1 / hour + threshold alert
Opening / contact	Event-driven

THE RIGHT INSTINCT

Unflagged interpolated data is **worse than a gap**: a gap says "I do not know"; while hidden interpolation asserts something false. The most useful cadence is the one matched to the dynamics of the phenomenon, not the most frequent one (which drains the battery).

GDPR, retention and sovereignty

STRAIGHT ANSWER

Governed data is data you can use with confidence, at scale and over time. Governance is not a layer bolted on afterwards: it is a property of AI-ready data.

- ✓ **Ownership:** every dataset has a business owner (quality, rules of use), distinct from the technical owner.
- ✓ **GDPR:** as soon as an asset indirectly tracks a person (a vehicle driven by an employee), an explicit purpose, transparency, minimisation, retention periods and data-subject rights all apply — enforced by your national data protection authority.
- ✓ **Retention:** keep only what is necessary, for as long as necessary — retention periods configurable by data type.

Sovereignty = economic security

Concentrating every position of a fleet draws the **operational map** of your business. Hosting in the **European Union** protects against extraterritorial legislation. Its corollary is **portability** — being able to recover all of your data and leave a supplier without obstruction.

WARNING SIGN

Personal data collected outside the legal framework (undocumented purpose, people not informed) is **unusable**: on the day of the project, you will not be entitled to use it. Compliance is prepared up front.

From collection to decision

STRAIGHT ANSWER

Understanding the end-to-end chain shows where quality is won and lost. Most of the value plays out in the first links: badly collected data is never rescued by a beautiful dashboard.

1

Collection

The tracker captures position and measurements. Quality = hardware, configuration (cadence, thresholds), real-world robustness.

2

Transport

The network carries the data (NB-IoT/LTE-M for autonomous assets, 4G for real time). Coverage gaps and offline storage play out here.

3

Ingestion & storage

Reception, deduplication, timestamping: this is where uniqueness and traceability are built.

4

Transformation & quality

Cleaning, normalisation, enrichment from the register, control of the 6 dimensions. The stage that produces AI-ready data.

5

Exposure

Dashboards, exports and above all machine interfaces (API, MQTT) for analytics tools and models.

The API: the front door

An **open, documented, two-way API** (JSON format) is the single most decisive technical criterion: without it, the data stays trapped inside the platform and every analysis becomes a manual export-import.

MQTT & your IT estate

MQTT (and webhooks) push events in near real time for reactive use cases. Value comes from cross-referencing with your ERP, TMS, GIS and CMMS, and from the **Power BI** export.

What clean data unlocks

STRAIGHT ANSWER

There is no point aiming for exotic models while the fundamental use cases go unserved. Four families create the most value — tackle them in order of the data maturity they require.

Use case	Typical gain (2025-26 studies)	Critical dimensions
1. Optimisation / surplus fleet	-10 to -15% surplus fleet	Completeness, consistency (register)
2. Anomaly and theft detection	Faster recovery, fewer losses	Consistency, timeliness, uniqueness
3. Predictive maintenance	-30 to -50% downtime, ROI ~10:1	Accuracy, completeness, traceability
4. Predictive management	More reliable investment decisions	All of them

THE ORDER MATTERS

Start with **surplus fleet** (a quick gain, undemanding), then **anomalies** (little history required). Those two steps fund what follows and build up the clean history that **predictive maintenance** needs. Starting with the most demanding use case is the number one cause of failure.

Sources: US Department of Energy; predictive maintenance studies 2025-2026; McKinsey. Indicative figures, to be validated on your own scope.

90 days to AI-ready data

STRAIGHT ANSWER

The worst strategy: clean everything before using anything. The right one is iterative and use-driven — pick a use case, prepare just enough, deliver, measure, extend.

1

Days 1-30 · Frame & found

Choose a priority use case and quantify what is at stake; build or clean the register; map the sources; establish the initial quality profile (6 dimensions).

2

Days 31-60 · Stabilise & measure

Fix the dimensions below threshold; put automatic controls in place (plausibility, deduplication, gap flagging, time); document the metadata; monitor quality continuously.

3

Days 61-90 · Expose & exploit

Expose through the API; enrich with a business source; launch the analytics pilot; measure against the success criterion, then decide whether to extend.

ADVICE

Organisations that succeed with AI projects are twice as likely to have **rethought their business process before choosing their tools** (McKinsey, 2025). Start with the business question and the data, not with the technology.

Checking that your fleet is ready

STRAIGHT ANSWER

Five recurring pitfalls ruin asset data projects. The checklist lets you verify, point by point, that a fleet is genuinely ready for AI.

The 5 pitfalls

- ✓ **Collect everything, use nothing** (no priority use case).
- ✓ **Underestimating the register** (the number one cause of failure).
- ✓ **Hiding uncertainty** (silently interpolating the gaps).
- ✓ **Deferring compliance** (personal data outside the legal framework = unusable).
- ✓ **Accepting vendor lock-in** (no export, no way out).

Checklist – is my fleet ready?

- ✓ Register: pivot identifier, closed taxonomy, key attributes, named owner
- ✓ Quality: 6 dimensions monitored, gaps flagged, timestamps normalised, deduplication in place
- ✓ Governance: documented purpose, people informed (GDPR), retention defined, EU hosting + portability
- ✓ Architecture: open API, real-time feed (MQTT), bridges into your IT estate (ERP/TMS/CMMS/Power BI), lineage

How TRAKmy delivers AI-ready data

STRAIGHT ANSWER

TRAKmy is designed to produce usable data, not just dots on a map. Every link in this guide has a native answer.

- ✓ **At the source:** rugged hardware (IP68/69K, IK09, up to 10 years) designed and manufactured in Europe. ULTRA N+ (1101) and its 5 sensors produce the usage series that predictive maintenance needs; COMPACT (1110) for tools; REACT (1203) for real time and mileage.
- ✓ **In transport:** NB-IoT (70+ countries), with offline storage that turns a permanent gap into a gap backfilled later.
- ✓ **In storage & quality:** the platform (Essential / Professional / Analytics), geofences, alerts, configurable retention (12 / 24 / 36 months). Geobeacon anchors for indoor-outdoor continuity.
- ✓ **In exposure:** API, MQTT (from Professional up), Power BI export (Analytics).
- ✓ **In governance:** data hosted in Europe (EU), European support, guaranteed portability.

IN SUMMARY

TRAKmy supplies the **reliable end-to-end chain**; the method in this guide gives you the register, the six dimensions and the roadmap. For a costed answer on your own fleet, ask for a quotation.

● APPENDIX · TO TAKE AWAY

The questions to ask your supplier

STRAIGHT ANSWER

Ask these at evaluation time: they separate platforms designed to be **integrated** from those designed only to be looked at.

Collection & hardware

- ✓ What is the real battery life at *my* cadence (not the catalogue figure)?
- ✓ Which sensors, at what configurable cadence?
- ✓ What documented field failure rate over several years?

Transport & integrity

- ✓ How long does the tracker store data offline, with guaranteed lossless upload?
- ✓ How are periods with no signal flagged?
- ✓ Are duplicate frames deduplicated automatically?

Exposure & integration

- ✓ Open, documented, two-way API? In what format?
- ✓ MQTT / webhooks for real time?
- ✓ Power BI export? Configurable retention? Full data recovery on exit?

Governance

- ✓ Where is my data physically hosted?
- ✓ How do I comply with the GDPR if I track vehicles driven by employees?



TAKE THE NEXT STEP

Let us turn your fleet into a reliable data source

Reliable hardware at the source, offline storage, configurable retention, API, MQTT and Power BI export, data hosted in Europe: TRAKmy supplies the AI-ready chain end to end. Ask for a demonstration and a quotation.

[Request a quote](#)

TRAKmy · La Station · 22 rue Pradier · 30000 Nîmes · France
contact@trakmy.fr · +33 4 30 00 57 25 · trakmy.fr
Data hosted in Europe (EU) · Hardware manufactured in Europe