From Data Quality Scores to Verifiable Evidence: Gaia-X Pattern for Trustworthy Data Products

16:25 – 16:55

- **Antonio J. Jara –** Libelium

In partnership with

gaia-x Hub Greece

LMS Laboratory for Manufacturing Systems & Automation

Data Quality in the Era of Data Spaces

Moving from static validation to portable semantic assurance.

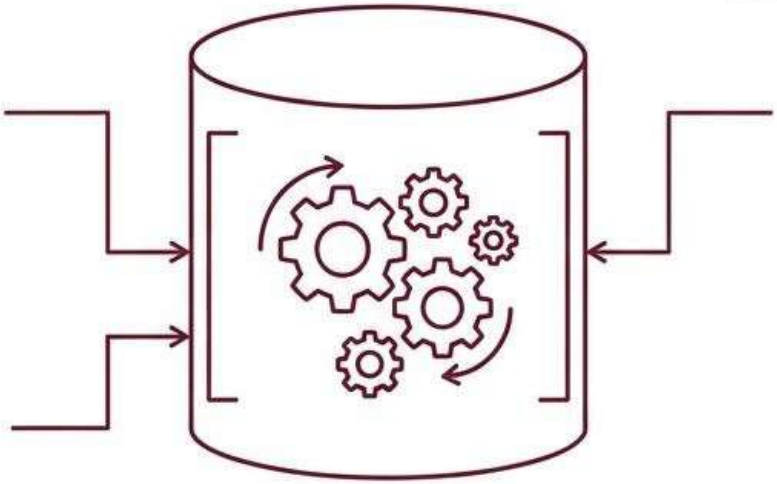
Dr. Antonio Jara, CTO
jara@libelium.com

Iris Cuevas, Head of Lab
i.cuevas@libelium.com

libelium

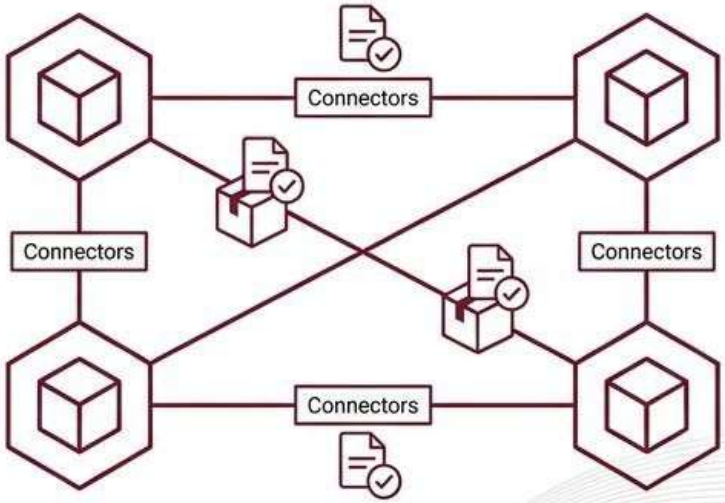
The Core Shift: From Validation to Assurance

Centralized Validation



The consumer controls the pipeline. Quality is a set of internal operational checks (nulls, types, freshness) over owned tables.

Federated Assurance

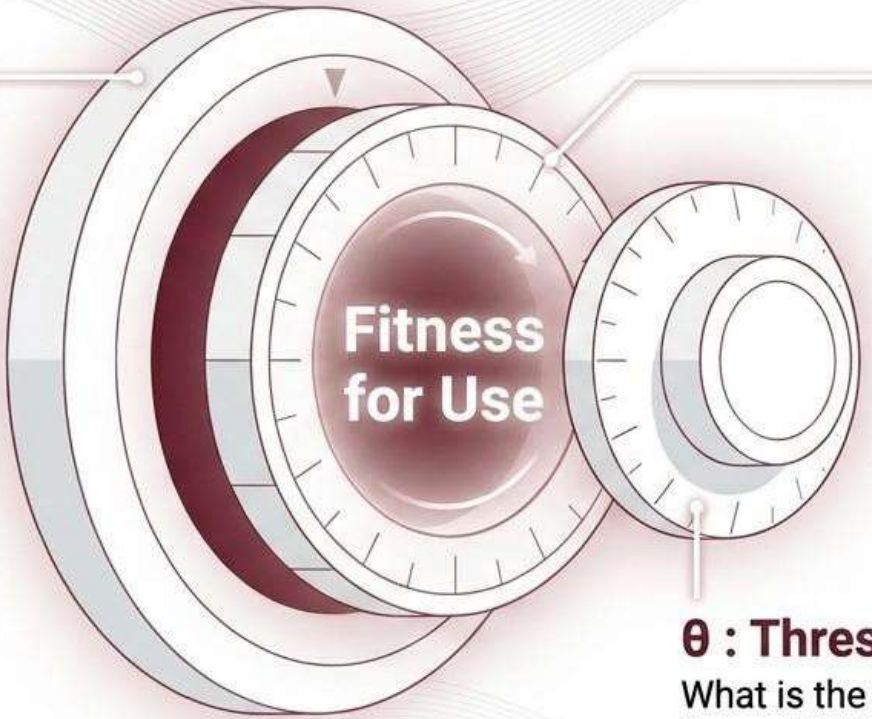


The consumer rarely controls the source, pipeline, or sensors. Quality becomes a form of transferable evidence that survives federation.

Unlocking Fitness for Use

d : Dimension

What characteristic are we measuring?
(e.g., Accuracy, Freshness)



m : Metric

What is the exact evaluation method? (e.g., Z-score outlier detection, max staleness)

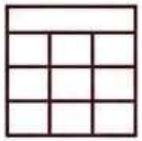
θ : Threshold

What is the acceptable decision criterion?
(e.g., > 0.95 confidence, < 2ms latency)

Data quality is not a universal scalar. It is the measurable degree to which a data product satisfies explicit requirements for a declared context.

Classical Dimensions Are No Longer Enough

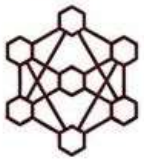
Classical DW Dimensions



- Accuracy
- Completeness
- Consistency
- Timeliness
- Uniqueness
- Validity

Necessary for table-level operational health.

Data-Space Requirements



- **Provenance**: Origins, lineage, and responsible agents.
- **Sensor Reliability**: Calibration age, security posture, tamper resistance.
- **Policy Conformance**: Alignment with Gaia-X/IDSA usage contracts.
- **Semantic Interoperability**: Compatibility with NGSI-LD schemas and vocabularies.
- **AI Readiness**: Drift metrics, bias evidence, spatial representativeness.

The EU AI Act makes this shift explicit: high-risk AI training datasets must carry auditable evidence of representativeness and relevance.

The Crisis of Decentralized Trust

When you do not own the sensor, how do you trust the data?

Local Validation is insufficient

Data crosses organizational borders

The need for Portable Assurance

Quality can no longer remain trapped in proprietary monitoring tools.
It must evolve into a transferable, machine-actionable credential.

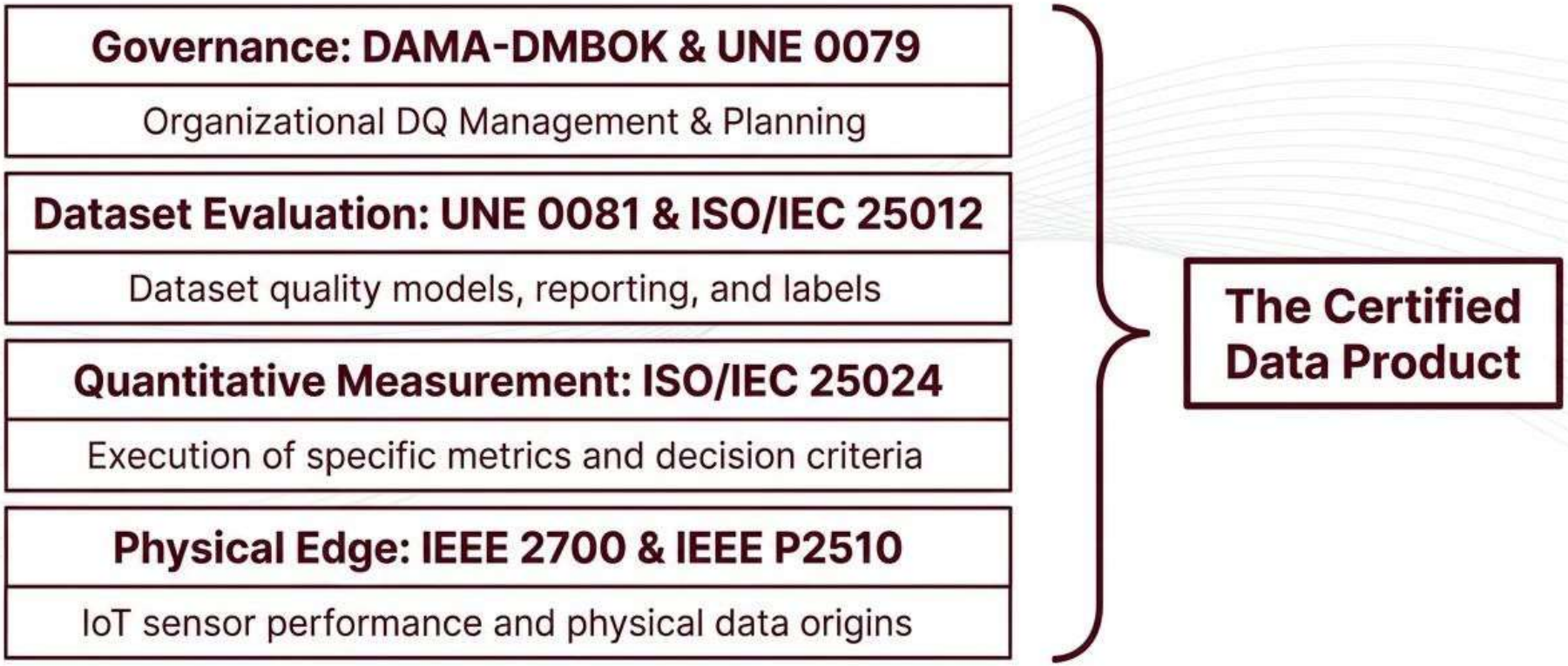
The Paradigm Shift in Data Quality

Redefining Fitness for Purpose

Dimension	Classical DQ	Data Space DQ
Ownership	Internal	Shared
Metric Location	Trapped in local dashboards	Attached to the data payload
Trust Model	Implicit (Organization-bound)	Verifiable (Contract-bound)
Target Output	Cleansed Tables	Certified Data Products

Decoding the Data Quality Standards

A Layered Architecture of Trust



#GaiaX #TechX

The Unified Quality Evidence Stack

Standard	Enterprise Scope	Dataset Scope	Sensor/IoT Scope	Semantic Exchange
DAMA & UNE 0079	Focuses purely on Enterprise Process (Planning, Stewardship, Governance).			
ISO/IEC 25012 & UNE 0081		Focuses strictly on Dataset Evaluation (Metrics, Thresholds, Reporting).		
IEEE 2700 & P2510			Anchors the Sensor Layer (Calibration, Hardware Accuracy, Digital Logical Units).	
W3C DQV, PROV & NGSI-LD				Powers the Semantic Exchange (Machine-actionable metadata, Trust chains).

Key Insight: These standards are **complementary**, not substitutable. IEEE P2510 guarantees the physical sensor, UNE 0081 grades the dataset, and **NGSI-LD carries the evidence**.

The Core Principle

Quality Evidence Must Travel With The Data

A dataset is only as valuable as the verifiable evidence attached to it.

Operationalizing Quality

Extending Smart Data Models (SDM)

The Goal

Make quality evidence queryable and machine-actionable.

The Method

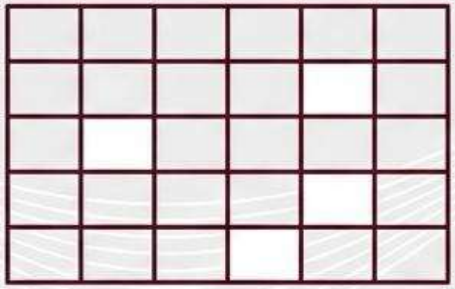
Libelium extends standard NGSI-LD payloads to include IEEE P2510 sensor indices and pipeline health metrics.

The Result

Two distinct representation patterns (Pattern A & Pattern B) that balance real-time performance with audit-grade evidence.

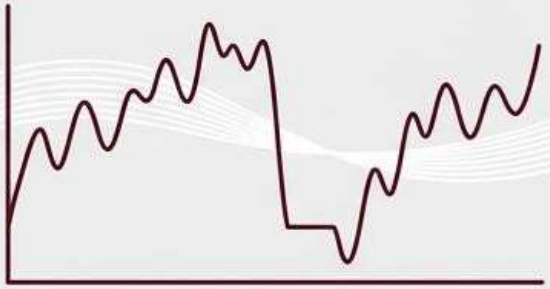
Context Changes the Shape of Quality

Data at Rest
(Batch)



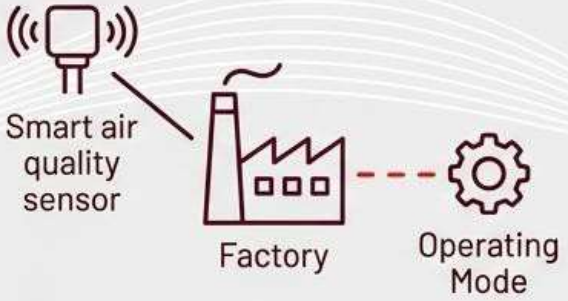
Measured via null ratios and absent attributes. Assessed via static dataset profiles.

Data in Motion
(Stream)



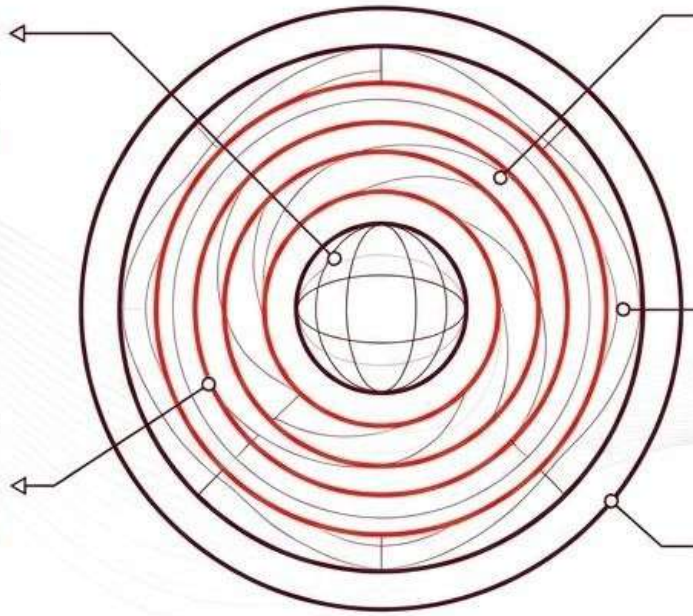
Measured via gap duration, missing bursts, and sampling regularity. Null ratios fail to capture critical event outages.

Data in Context
(Semantic)



Measured by contextual relevance. Data may be mathematically complete but unusable if asset hierarchy, operating mode, or semantic state are missing.

The Natural Unit of Trust



Core: Raw IoT Telemetry
(The actual values and physical readings)

Smart Data Models
(Entity types, relationships, units)

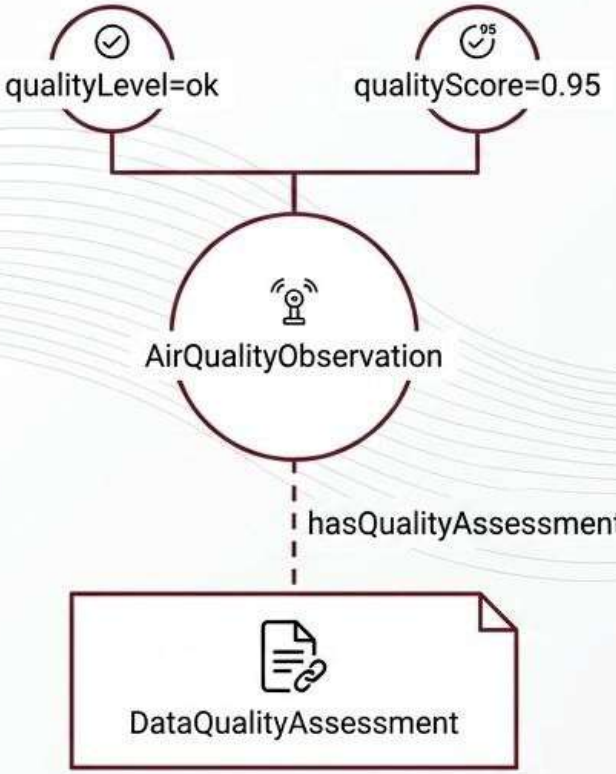
W3C PROV
(Provenance of sensors, transformations, and agents)

ODRL Policies
(Permissions, data-space usage agreements)

Quality Passports
(DQV metrics, Verifiable Credentials, IEEE P2510 DLUs)

Instead of publishing raw files, data space participants publish this bundle.
Quality evidence travels with the data.

Semantic Exchange via NGSI-LD



Property-level metadata enables high-speed operational routing and context broker filtering.

Linked entities preserve audit-grade evidence, rule execution histories, evaluator identities, and verifiable Gaia-X credentials.

Takeaway: This dual pattern preserves both runtime performance and regulatory auditability.

Composable Tooling: Execution vs. Representation

Execution Layer

Great Expectations, Deequ, dbt, Databricks, Monte Carlo.

Run checks, batch profiles, stream monitors, anomaly detection. Results remain isolated inside proprietary tool boundaries.

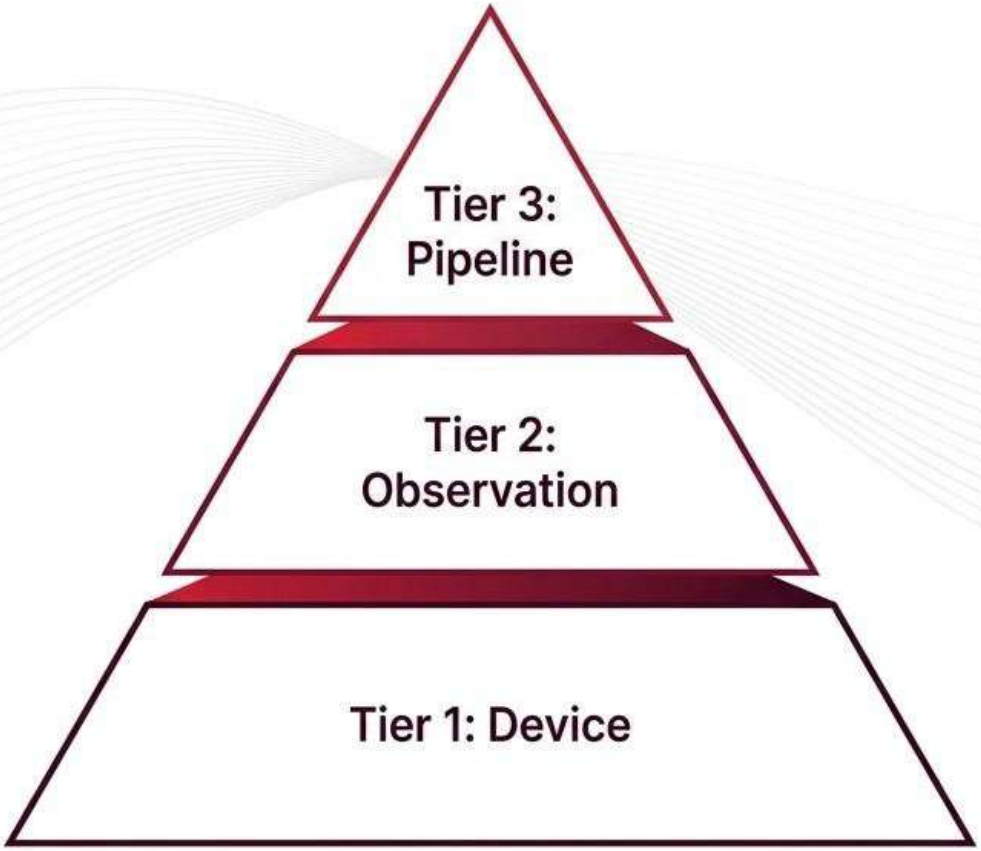
Semantic Translation Layer

Representation Layer

W3C DQV, NGSI-LD, Gaia-X Verifiable Credentials.

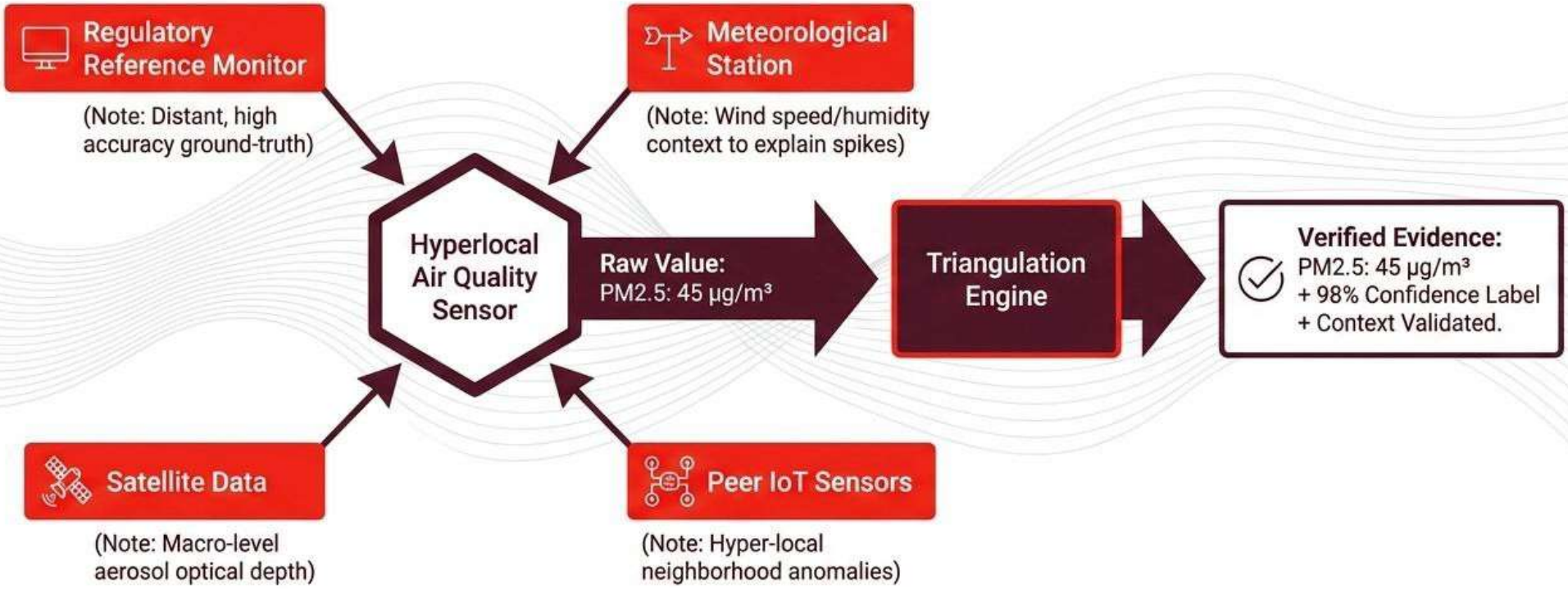
Execution outputs are normalized into a common ontology, attached to data products, and made usable by data-space connectors.

The Air Quality Evidence Hierarchy



Parameters	Standard
Sampling cadence regularity, gap duration, watermark lateness, data-loss windows, broker freshness.	DAMA / UNE 0079 stream metrics
Measurement plausibility, contextual anomaly status, uncertainty intervals, imputation status.	ISO 25024 applied to streams
Calibration date/method, hardware precision bounds, firmware version, tamper resistance, power state.	IEEE P2510

Context Validation Through Triangulation



Cross-validation transforms internal sensor assertions into external marketplace evidence.

Bridging the Gap at the Edge

Integrating Physical IoT Quality with IEEE P2510

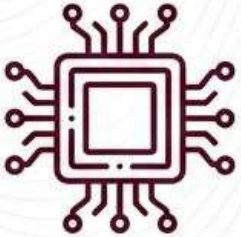
The Blind Spot

Classical DQ ignores sensor physics.

The Physical Reality

Calibration status, tamper resistance, energy levels, and firmware security.

The Digital Logical Unit (DLU)



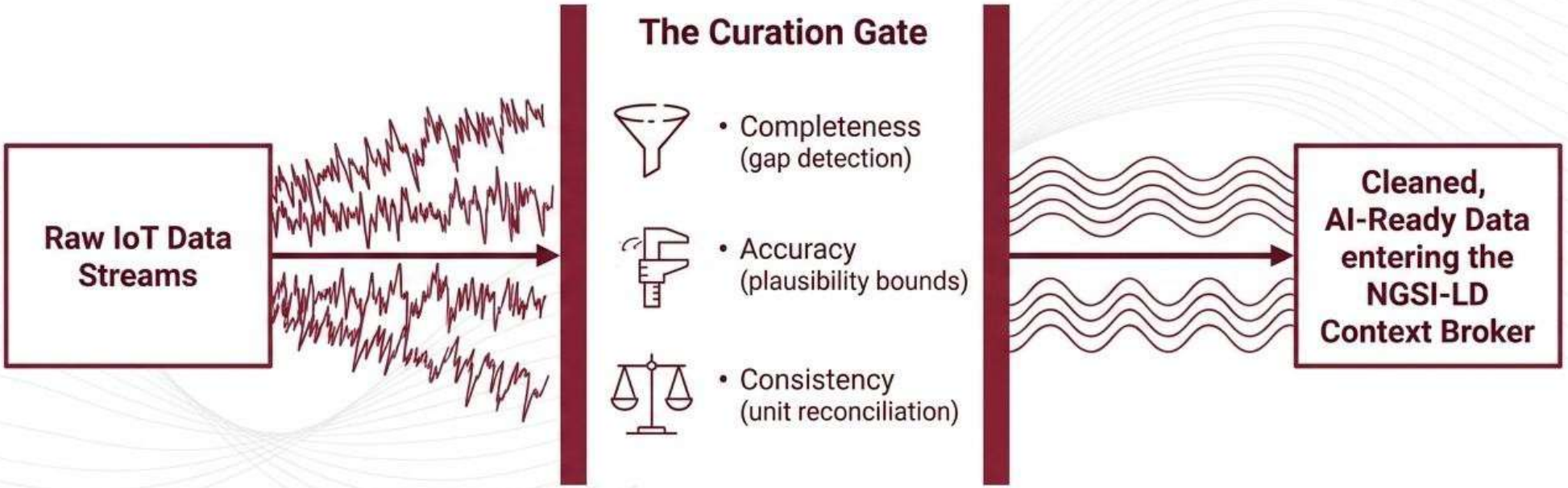
IEEE P2510 translates physical sensor reliability into a queryable digital index.

High-quality AI cannot be trained on uncalibrated, unverified hardware.
The DLU bridges physical operations with digital analytics.

#GaiaX #TechX

The Curation Firewall

Protecting the Context Broker



Core Rule: No data enters the context broker without passing a curation gate and receiving metadata tagging.

Pattern A: Fast Filtering

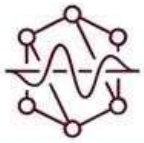
Sub-properties for Machine Interpretability

Raw Payload

```
{
  "temperature": 17.2
}
```

Enriched NGSI-LD Payload (Insight Data Enricher)

```
{
  "temperature": {
    "value": 17.2,
    "qualityLevel": "validated",
    "qualityScore": 0.92,
    "qualitySource": "IEEE_P2510_Rule_X"
  }
}
```



Strategic Purpose: Enables real-time digital twins to instantly filter out suspect or outlier data without complex external queries. Preserves backward compatibility.

Pattern B: Audit-Grade Evidence

The DataQualityAssessment Entity

Strategic Purpose: Essential for Data Space certification, ISO 17029 compliance, and detailed execution logs.

Continuous Pipeline Observability

Monitoring Freshness, Volume, and AI-Drift

Databricks Intelligence

Automated freshness and completeness tracking using historical commit patterns.

Great Expectations

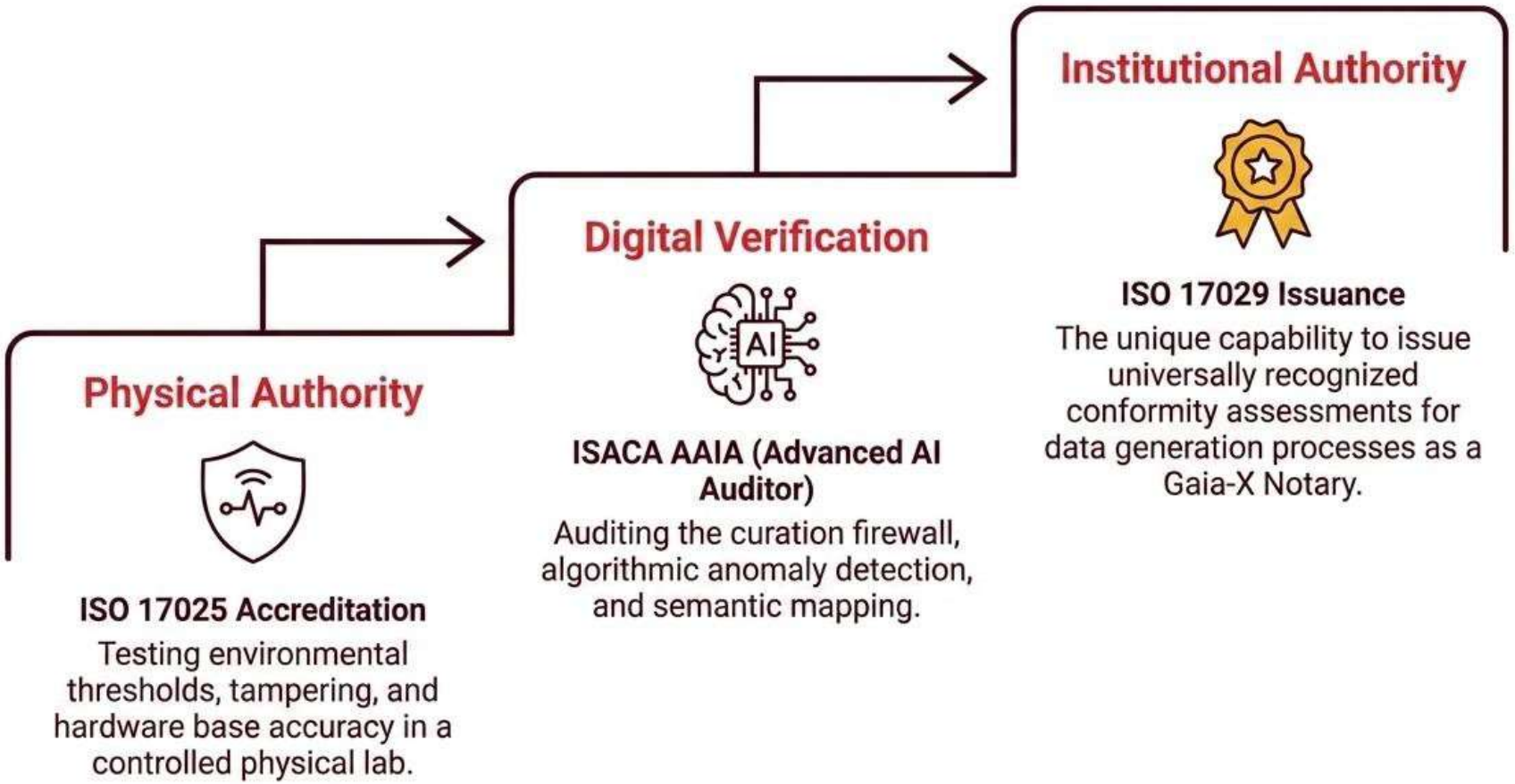
Rule-based validation, distribution checks, and schema enforcement.

Feedback Loop

Pipeline anomalies are published back into the NGSI-LD semantic graph, altering the DataQualityAssessment.

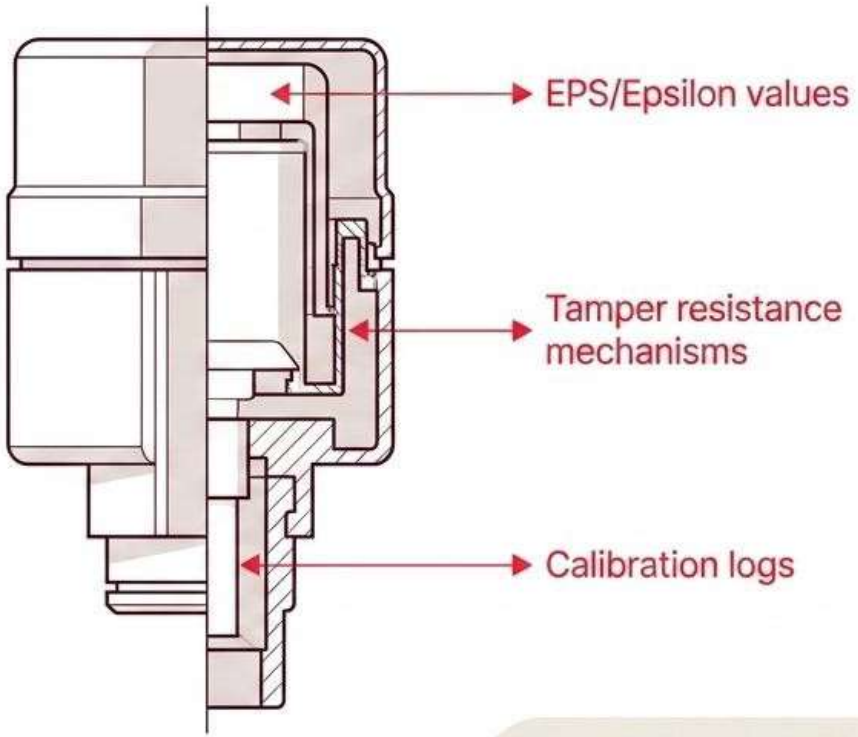
Key Concept: Continuous evidence ensures downstream AI models are protected from stale or drifting datasets.

The Libelium Trust Triad



#GaiaX #TechX

Step 1: Anchoring Trust in Hardware Truth



IEEE P2510

Defines global benchmarks for IoT data quality indices (Accuracy, Reliability, Feasibility).

CEN 17660

Specific normative benchmarks for air quality measurement and environmental constraints.

Trust begins at the edge. If the physical signal is compromised, downstream AI logic is invalid.

#GaiaX #TechX

Operationalizing Quality in NGSI-LD

The Goal

Make quality evidence queryable and machine-actionable across federated marketplaces.

The Method

Extending standard NGSI-LD Smart Data Models (SDM) to embed IEEE P2510 sensor indices directly into the data product.

The Result

A standardized payload carrying both the measurement and the cryptographic proof of its trustworthiness.

Smart Data Models: Representation Patterns

Pattern A: Operational Speed

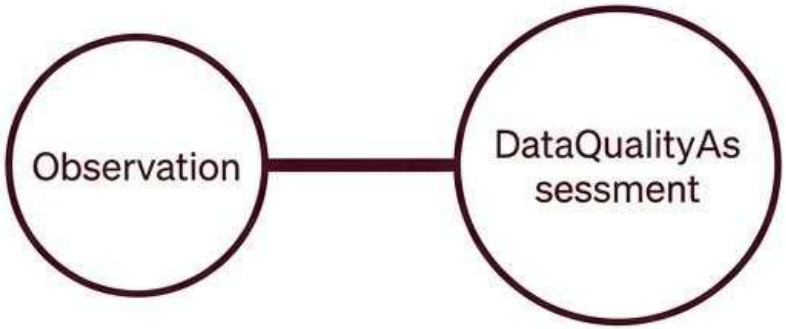
Sub-properties

```
{
  "temperature": 24.5,
  "qualityLevel": "valid"
}
```

Fast filtering. Query all measurements where qualityLevel = valid. Human and machine interpretable.

Pattern B: Institutional Audit

Related Metadata Entities



Audit trails and certification evidence. Stores detailed rule execution logs, DIH assessments, and hardware calibration records.

The Data Quality Assessment Ecosystem

From Standards to Smart Data Models

1. DATA QUALITY STANDARDS & VALIDATION (INPUTS)

IEEE P2510
(Sensor Params: EPS, Security, Calibration)

DAMA (DMBOK DQ Framework) & **ISO 8000**
(Global DQ Standard)

Defines rigorous global benchmarks & metadata indices.

2. TYPES OF ANALYSIS (ASSESSMENT CORE)

ACCURACY
Target, market and medicallzation

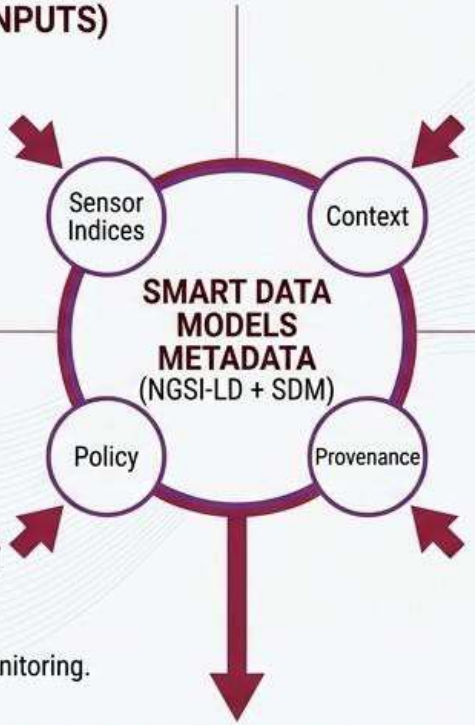
COMPLETENESS
Puzzle to ansors & completeness

RELIABILITY
Anchor descriptions & on-moctrennate

TIMELINESS
Data timeness and timesliminiute

RELEVANCE
Filter to bowi and anel relevance

CONSISTENCY
Linked chain & chontn analyass



3. DATA QUALITY MONITORS & CURATION (OPERATIONALIZATION)

CURATION FIREWALL
(Gate, Filter)

AI-DRIVEN DETECTION
(Brain chip)

PIPELINE/DATASET MONITORS
(Dashboard)

Pre-broker validation, ML-based checks, automated freshness monitoring.

4. TYPES OF ANOMALIES & DRIFTING (DETECTED ISSUES)

OUTLIERS
Red dot pranchmards & deciror of outliers

MISSING DATA
Empty box a resontans trust of missing data

DRIFTING
Trend line calls after an drifting (Trend line)

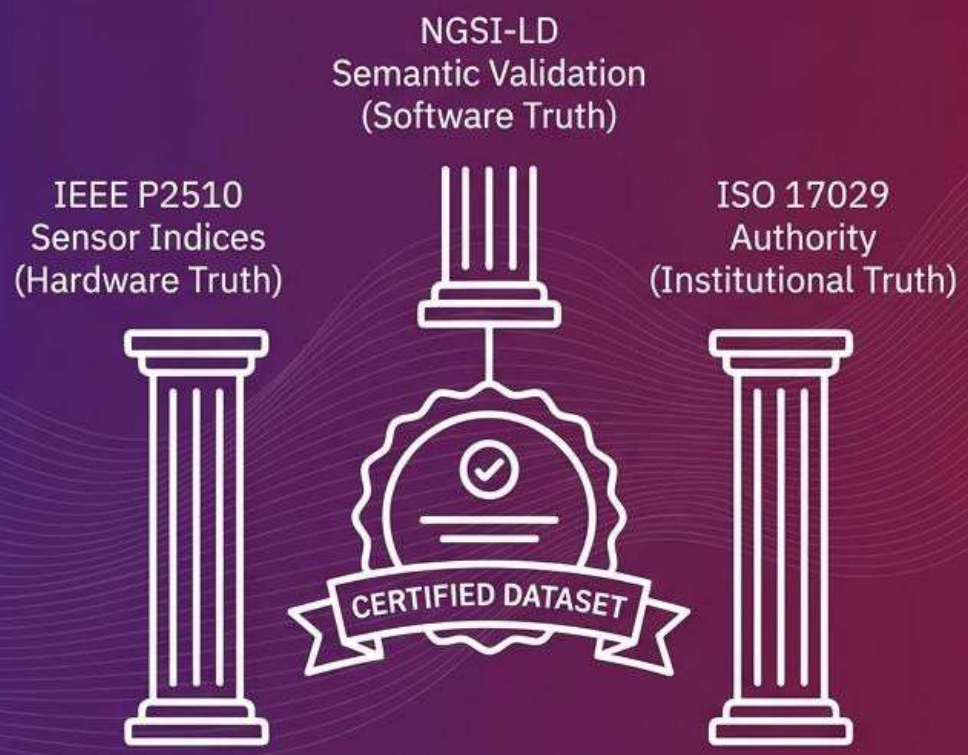
INCONSISTENCIES
Inconsistencies (Conflicting arrows)

TRUSTED DATA PRODUCTS & CERTIFICATION (OUTCOME)

DATA QUALITY LABEL
(DIH Certified)

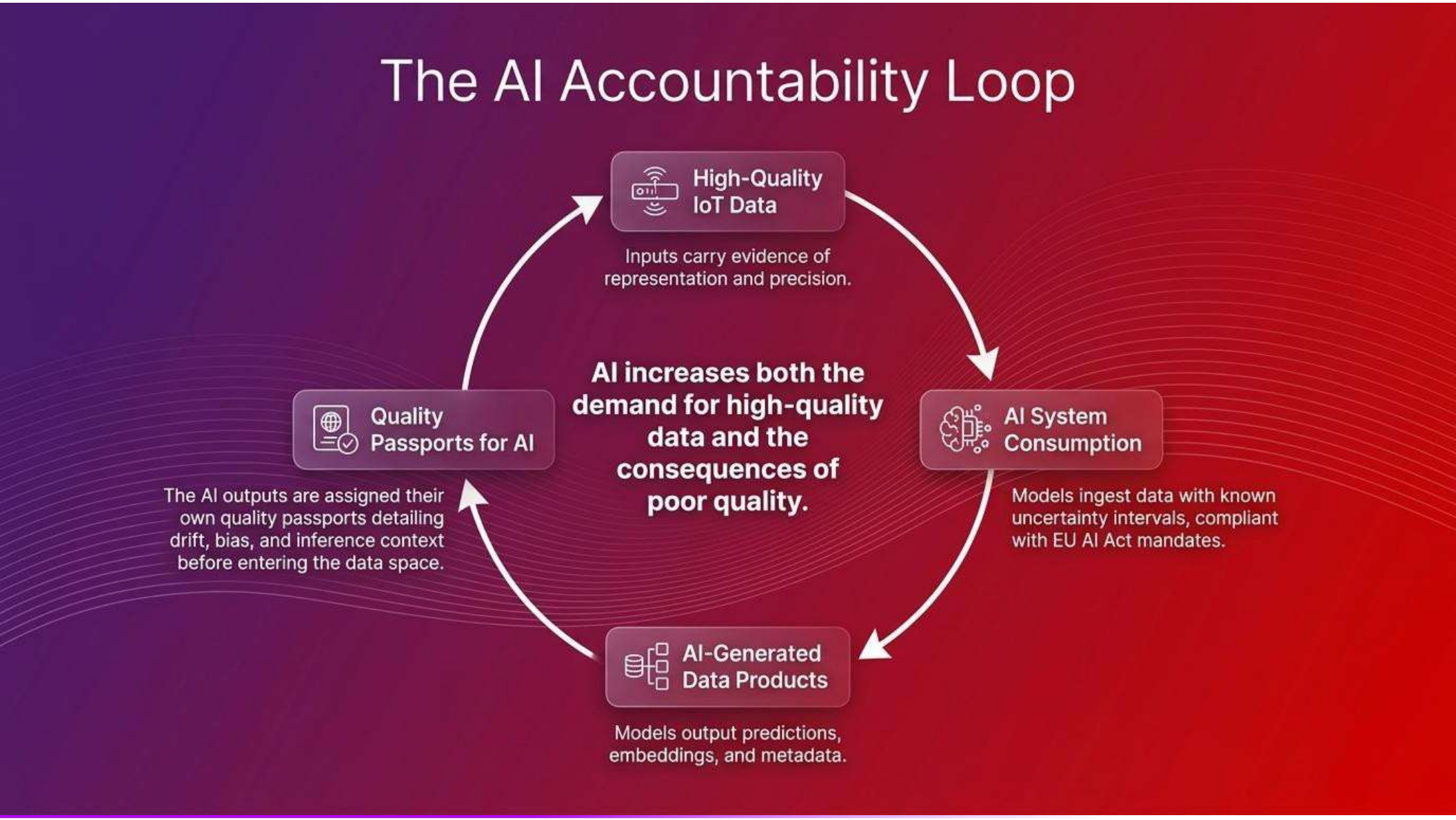
Auditable evidence, increased trust, reduced integration costs. Enables reliable Digital Twins & Industry 4.0.

The Output: The Verifiable Data Quality Label



- ✓ Issued directly as a Verifiable Credential (VC).
- ✓ Drastically reduces integration costs for SME data consumers.
- ✓ Pre-validates AI-training datasets for EU AI Act compliance.

The AI Accountability Loop



The Marketplace Trust Deficit

In a sovereign Data Space, participants do not trust metrics; they trust authorities

Self-assessed metrics fail in federated markets.

Data must carry mathematical guarantees.

Guarantees must be backed by a recognized Trust Anchor.

From Engineering Metric to Market Property

Pricing & Discoverability

Datasets with documented provenance, explicit quality vectors, and verifiable labels command premium marketplace value and higher catalog visibility.

Integration Friction

Smart Data Models and semantic quality layers drastically reduce the due-diligence and schema-mapping costs for data consumers.

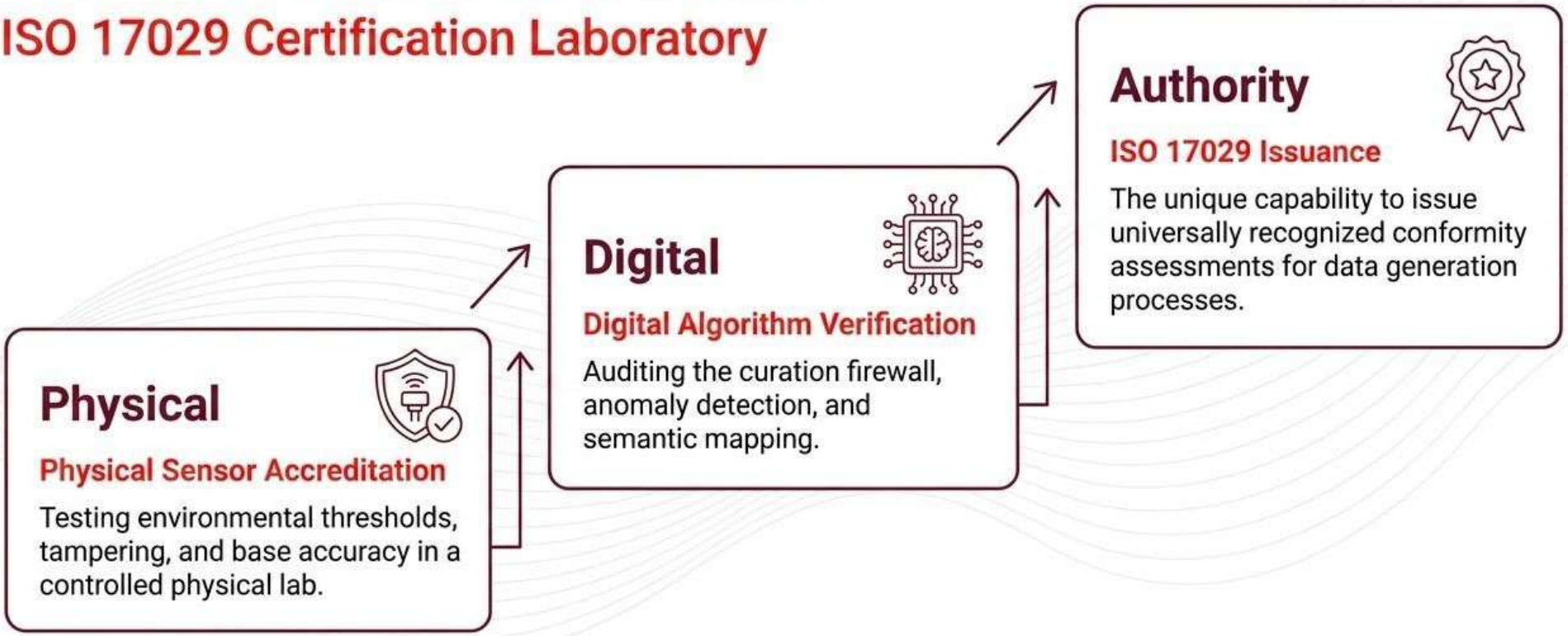
Liability & Risk

Transparent uncertainty bounds and explicit usage exclusions protect data providers from contractual risk and regulatory non-compliance.

Quality is a differentiator that allows providers to monetize trust.

The Libelium Trust Anchor

ISO 17029 Certification Laboratory

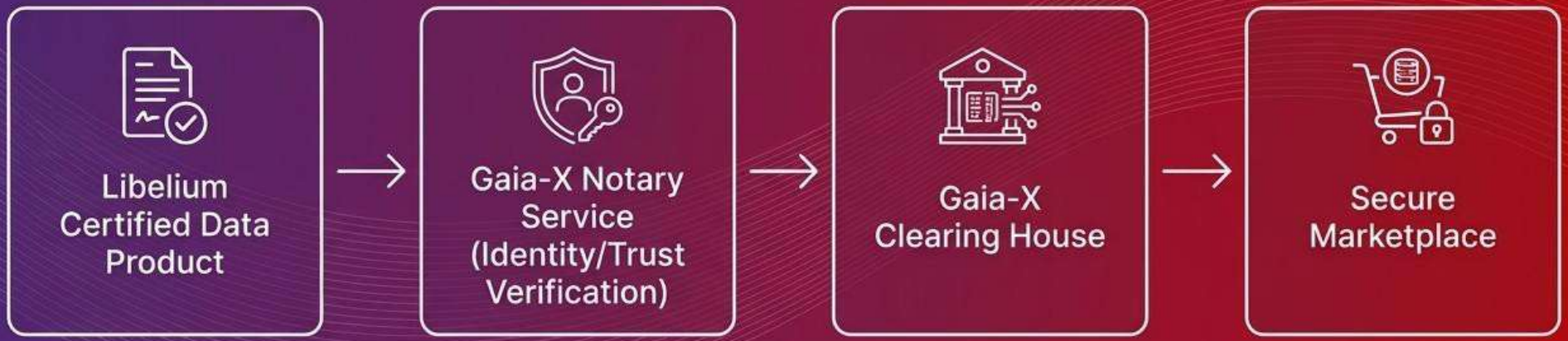


Physical lab accreditation translates directly to digital trust in Gaia-X.

#GaiaX #TechX

Building the Gaia-X Trust Ecosystem

Frictionless, Monetizable Data Sharing



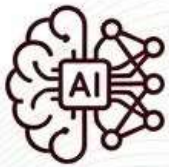
Libelium's quality labels integrate natively with Gaia-X Clearing Houses, acting as verifiable credentials that unlock frictionless data negotiation and premium monetization.

The Business Impact of Certified Data

Value Creation for Industry 4.0 and Smart Cities

Reduced Friction

Standardized payloads cut data integration costs. Due diligence is pre-completed by the Libelium label.



Trusted AI

AI models trained on certified data comply with emerging regulations and reduce hallucination/drift risks.

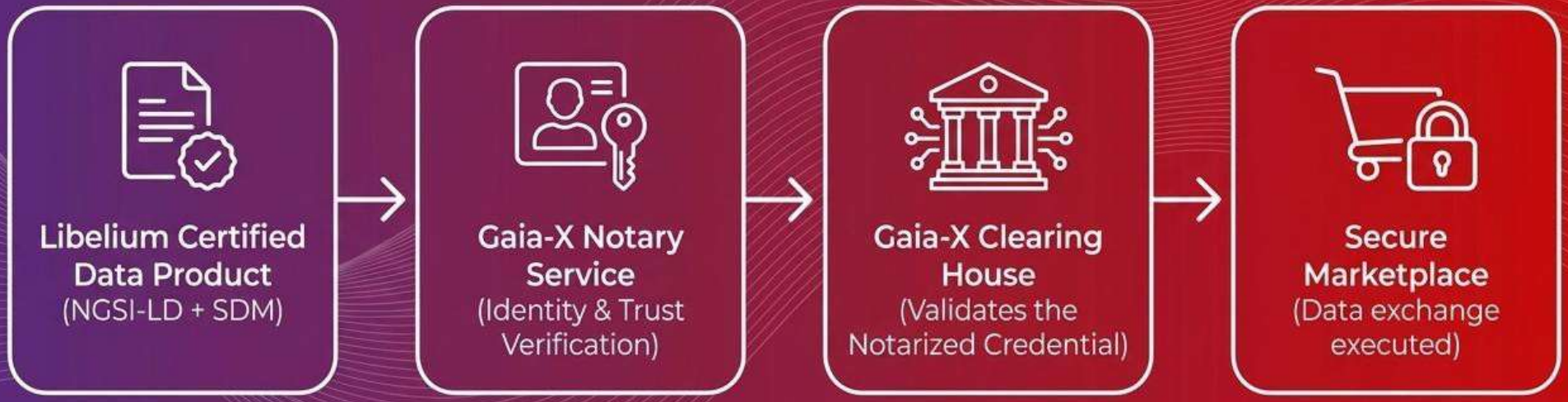
Premium Monetization

Certified datasets command premium pricing in data spaces due to guaranteed Service Level Objectives (SLOs).

The Paradigm Shift in Data Quality

Dimensions	Self-Assessed Metrics	Verifiable Evidence (Libelium)
Source of Truth	Internal dashboards	Cryptographic proofs
Governance	Best-effort policies	Institutional auditing (ISO 17029)
Output Format	Proprietary logs	Standardized Verifiable Credentials (VCs)
Market Interoperability	Siloed within platforms	Natively queryable via NGSI-LD
AI Readiness	Requires manual due diligence	Ready for EU AI Act compliance

Frictionless Sharing in the Gaia-X Gaia-X Ecosystem

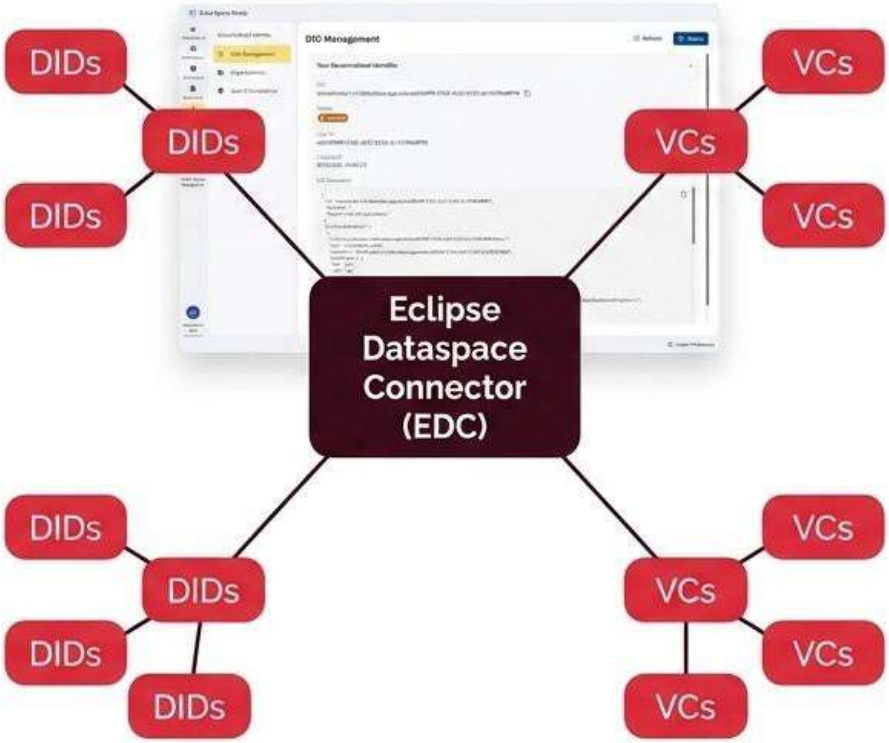


Libelium's quality labels integrate natively with Gaia-X Clearing Houses, acting as the cryptographic key that unlocks frictionless data negotiation.

Powering the EU LDT Toolbox

LDT CitiVERSE EDIC

The Identity Fabric



The Impact

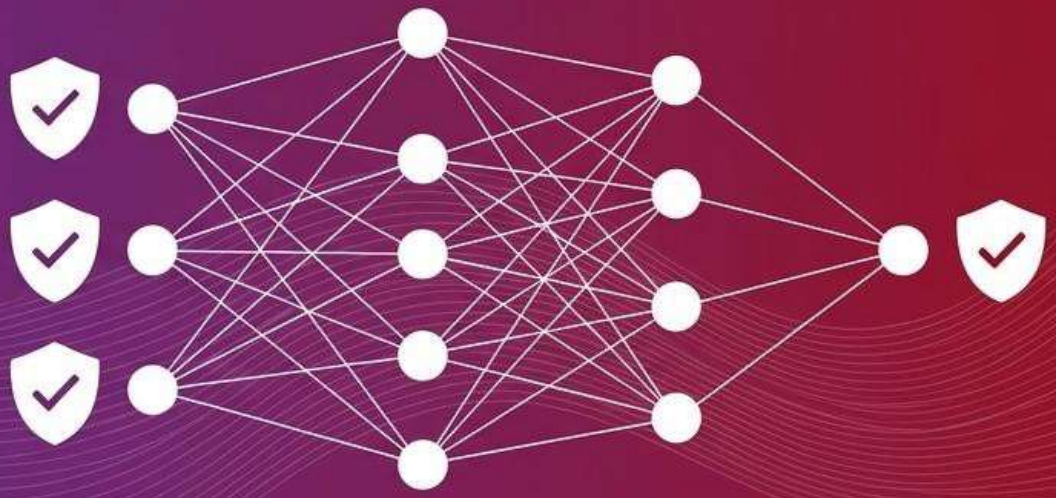
Decentralized identity ownership ensures city sovereignty over Smart City data.

Libelium's Notary services validate organizational roles and data compliance across the federation.

Enables seamless, horizontally scalable data exchange between previously siloed municipal platforms.

The Citcom.ai TEF Imperative: Trusted AI

CitCom AI



Verifiable Evidence In, Compliant AI Out

The EU AI Act mandates strict provenance and quality tracking for training data.

Self-assessed datasets expose developers to unacceptable regulatory and hallucination risks.

Libelium's Data Quality Label provides the definitive audit trail required for Citcom.ai's advanced machine learning pipelines.

The Business Impact of Certified Data Products

Reduced Friction

Standardized payloads and SDM patterns cut data integration costs. Due diligence is pre-completed by the Libelium label.

Trusted AI

Models trained on certified data comply with emerging regulations and dramatically reduce drift and hallucination risks.

Premium Monetization

Certified datasets command premium pricing in federated marketplaces due to cryptographically guaranteed Service Level Objectives (SLOs).


Trust is not assumed. It is verified.
Standardize your IoT Data Products with Libelium.
Discover the ISO 17029 Data Quality Laboratory.