

Detailed Reconstruction Cost Breakdown

(Public Cloud Project Deletion – Scientific R&D Dataset & ML Stack)

1. Field Re-Sampling & Logistics

Scope: Re-establish microscopy sampling campaigns across 40–50 wastewater treatment plants.

Activities

- Site re-engagement & permissions
- Scheduling multiple seasonal sampling cycles
- On-site sample collection (influent, aeration basins, clarifiers, sludge lines)
- Sample preservation & transport
- Operator coordination & compliance

Cost Drivers

Item	Unit	Qty	India (€)	Global (€)
Site visit (travel + ops)	per plant	40–50	1,500–2,000	4,000–6,000
Multi-cycle sampling	cycles	2–3	1,200–1,800	3,000–5,000
Logistics & consumables	per plant	40–50	600–900	1,500–2,500

Subtotal

- **India:** €70,000 – €120,000
- **Global:** €220,000 – €380,000

2. Microscopy Data Regeneration (Raw Corpus)

Scope: Rebuilding ~50–55 million raw microscopy images.

Activities

- Sample preparation (staining, fixation)
- Microscopy operation (brightfield / phase contrast)
- High-resolution image capture
- QC & artifact rejection
- Archival structuring

Cost Drivers

Item	Unit	Qty	India (€)	Global (€)
Microscope operator	per FTE/year	6–10	8,000–12,000	40,000–60,000
Lab technicians	per FTE/year	4–6	6,000–9,000	35,000–55,000
Equipment depreciation	fixed	—	15,000–25,000	40,000–70,000
Consumables	per image batch	—	20,000–35,000	60,000–90,000

Subtotal

- **India:** €90,000 – €140,000
- **Global:** €260,000 – €430,000

3. Expert Annotation & Scientific Labeling

Scope: Re-annotating ~200,000 expert-labeled images.

Activities

- Microbial morphotype identification
- Functional group classification
- Cross-review & consensus validation
- Ontology mapping & dataset consistency checks

Cost Drivers

Item	Unit	Qty	India (€)	Global (€)
Domain experts	per image	200k	0.40–0.70	1.5–3.0
Senior reviewer QA	% of dataset	20–30%	10,000–18,000	30,000–60,000

Subtotal

- **India:** €90,000 – €140,000
- **Global:** €300,000 – €600,000

4. Soft-Sensor & Time-Series Data Reconstruction

Scope: ~12 million datapoints tied to plant operations.

Activities

- SCADA/PLC data re-extraction
- Historical signal alignment
- Sensor normalization & drift correction
- Feature engineering & interpolation
- Validation against operational benchmarks

Cost Drivers

Item	Unit	Qty	India (€)	Global (€)
Control systems engineer	FTE/year	2–3	12,000–18,000	70,000–100,000
Data engineer	FTE/year	2–3	15,000–22,000	80,000–120,000
Integration tooling	fixed	—	8,000–12,000	20,000–35,000

Subtotal

- **India:** €55,000 – €85,000
- **Global:** €180,000 – €300,000

5. ML Model Redevelopment & Validation

Scope: Loss of production-grade models, embeddings, experiments, and weights.

Activities

- Dataset re-curation
- Model retraining (multiple families)
- Embedding regeneration
- Hyperparameter tuning
- Regression & performance validation

Cost Drivers

Item	Unit	Qty	India (€)	Global (€)
ML engineers	FTE/year	3–4	18,000–28,000	90,000–140,000
Research scientist	FTE/year	1–2	22,000–35,000	110,000–160,000
Experiment tooling	fixed	—	8,000–15,000	25,000–45,000

Subtotal

- **India:** €80,000 – €130,000
- **Global:** €260,000 – €420,000

6. Compute Infrastructure (Training & Storage)

Scope: GPU compute + interim storage during rebuild.

Cost Drivers

Item	Unit	Qty	India (€)	Global (€)
GPU compute	months	6–9	25,000–45,000	70,000–120,000
Temporary object storage	TB-months	60–80	5,000–8,000	15,000–25,000

Subtotal

- **India:** €30,000 – €55,000
- **Global:** €85,000 – €145,000

7. Program Management, QA & Opportunity Cost

Scope: Coordination, delays, and lost velocity.

Activities

- Program oversight
- QA audits & documentation
- Research roadmap re-alignment
- Lost opportunity cost (pipeline delay)

Subtotal

- **India:** €35,000 – €55,000
- **Global:** €120,000 – €200,000

Final Aggregated Reconstruction Cost

Scenario	Total Range
India-Optimised Execution	€450,000 – €650,000
Global Market Execution	€1.5M – €2.9M

Critical Clarification (For Claims Team)

- The **5–10% local data** does **not materially reduce** any of the above line items.
- The dominant costs are **human expertise, time, and re-execution**, not storage volume.
- Model weights and embeddings are **non-derivable** without original training corpora.