

# Cithorum Knowledge Graph Explainer

## ***One sentence***

Jam makes the data cheaper to store, move, and restore. The Cithorum Knowledge Graph turns the Jam-compressed data plane into an operating map: customers, workloads, evidence, incidents, costs, decisions, and outcomes connected in one queryable system. The first KG runs the data centre itself.

## ***Why the KG exists***

Most AI products sit on top of fragmented tools. They can answer local questions, but they do not know the whole operating picture: which customer owns which workload, which restore was verified, which dataset drove which invoice, which incident changed the SLA, or which evidence supports the answer.

The Cithorum KG is the memory layer above Jam. It keeps the relationships live while Jam handles the economics underneath. Together they form one product: a managed sovereign cloud that compresses the bytes and indexes the operating reality of running them.

## ***The first KG is the data centre***

The first version is **Data-Centre Ops KG** — the operating graph of Cithorum’s own 1 PB Jam Proof Cloud. It is not speculative, it is not a vertical pilot, and it is not waiting for a design partner to start producing value. It runs the proof environment.

Every customer workload Jam touches creates structured telemetry:

- customer
- bucket
- dataset
- raw size
- Jam size
- ratio
- restore status
- CPU and throughput
- invoice basis
- node health
- incident history
- evidence artefacts
- capacity headroom
- energy draw
- cost per TB

The KG turns that telemetry into an operating graph. The first buyer is internal: Cithorum uses it to run the proof environment, generate monthly customer reports, prove restore drills, and produce the billing exports each tenant receives. The second buyer is external: every Cithorum

customer gets a slice of their own operating graph as part of the managed service. The third buyer is other data-centre operators who license the same engine to run their own.

## **Core architecture**

Customer data

- > Jam data plane
- > telemetry and metering
- > KG Core
- > data-centre operating products

## **Jam data plane**

Compression, indexing, cryptographic envelope, deterministic restore, and per-customer storage isolation.

## **Telemetry and metering**

Events emitted by Jam: size, ratio, restore result, throughput, CPU, workload type, customer, bucket, billing basis, capacity headroom, energy draw, incident, recovery.

## **KG Core**

Typed entities, typed relationships, provenance, permission checks, query layer, and AI-accessible summaries. The substrate is intentionally domain-agnostic: the same core that runs Data-Centre Ops can be re-aimed at any other relational domain when paid demand justifies it.

## **Data-centre operating products**

Customer dashboards, monthly benchmark reports, restore-proof drills, billing exports, capacity forecasts, incident reviews, and AI assistants grounded in graph evidence. These ship as part of the Cithorum Cloud managed service — not as separate products.

## ***What it can answer (data-centre operations)***

- Which customers are consuming the most raw capacity?
- Which workloads produce the strongest Jam savings?
- Which datasets have not had a recent restore proof?
- Which node, disk, queue, or customer bucket caused an incident?
- Which pilot should convert to paid based on measured savings?
- Which benchmark report supports this invoice?
- Which customer proof can be shown publicly, privately, or under NDA?
- Which partner route created which lead?
- Which workloads map to the next infrastructure raise?

This is the queryable operating reality of running a 1 PB sovereign cloud — graphed, audited, restorable, and exportable on demand. It is what hyperscalers expose as separate products (CloudWatch + Cost Explorer + Audit Trail + Neptune), bundled into one in-platform query layer with no egress fees and no separate bill.

## ***Why this belongs with Jam***

Jam creates the telemetry that most KG platforms lack: byte-level economics, restore proof, evidence trails, and customer workload structure. The KG makes that telemetry useful. Together they create the wedge:

- **Jam proves measurable savings.** 3–8× on production workloads, up to 100× on backup tiers.
- **The Ops KG proves operational intelligence.** Every restore is verified; every invoice is sourced; every incident is graphed against root cause.
- **The substrate compounds.** Jam telemetry is the most structured operational data any storage system produces. Each customer onboarded enriches the graph schema for the next.

## ***Build order***

### **Phase 1 — Jam telemetry (live)**

The Jam path is shipping: upload, process, compress, restore, meter, report. M2M TechConnect runs on it today at \$7K/month MRR.

### **Phase 2 — Data-Centre Ops KG (in progress)**

Graph the proof environment: customers, buckets, workloads, nodes, restores, incidents, costs, energy, evidence. Internal-first; external slices land as customer dashboards and benchmark reports.

### **Phase 3 — Customer-facing operating products**

Expose benchmark reports, restore proofs, billing exports, capacity forecasts, and workload history to every tenant of the Cithorum Cloud.

### **Phase 4 — Re-license the engine**

Other data-centre operators run the same Ops KG over their own Jam telemetry. Same substrate, different deployment. This is the natural commercial extension before any vertical work.

## ***Where the KG extends next***

The same core schema — typed entities, typed relationships, provenance, permissions, AI tooling — re-aims at vertical domains when paid design-partner demand exists. Each vertical is a different schema and a different buyer, but the core is shared.

- **Workspace KG** — business memory across tools, documents, decisions, and customers. Mid-market revenue ops.
- **Law KG** — private matter memory, precedent graph, audit-trailed AI over legal work. Law firms and in-house legal.
- **Medical-Sequencing KG** — sample, FASTQ, variant, evidence, and audit trail connected for genomics labs and clinical research.

These are real product directions, not speculative ones — Cithorum has a confirmed first design partner for Workspace KG and clinical-science routes through the Suraj S Naik / Chirag Labs network for Medical-Sequencing. They will be developed sequentially, design-partner-led, one at a time. Until then, Data-Centre Ops KG is the live product and the canonical proof of the substrate.

### ***Investor-safe positioning***

The KG is not a separate science project. It is how Cithorum's data-centre service becomes a platform. Jam lowers the cost of the data plane; the KG captures the memory, evidence, and workflows that compound on top of it. The proof environment is already the first product. The verticals are the next product. The core is the same.