

Data Provenance & Workflow Traceability — Hands-on

Make every result reproducible and auditable. This module teaches end-to-end data provenance and workflow traceability for omics, imaging, and clinical analytics—spanning standards (W3C PROV, CWLProv, RO-Crate), run metadata capture in Snakemake/Nextflow/Galaxy, dataset versioning (DVC/Pachyderm), experiment tracking (MLflow/W&B), integrity controls, and FAIR packaging with persistent identifiers.

Data Provenance & Workflow Traceability Systems

[Help Desk](#) · [WhatsApp](#)

Session Index

[Session 1 — Provenance Standards & Lineage Graphs](#) [Session 2 — Versioned Data & Experiment Tracking](#) [Session 3 — Pipeline Run Metadata & Integrity](#) [Session 4 — Mini Capstone: FAIR, RO-Crate & Report](#)

Session 1

Fee: Rs 16,900 [Apply Now](#)

Provenance Standards & Lineage Graphs

Concepts & standards

[W3C PROV](#) [CWLProv](#) [RO-Crate](#)

Lineage modeling

[entities/activities/agents](#) [graph stores \(Neo4j\)](#)
[SPARQL/JSON-LD](#)

Provenance queries & views

impact analysis explainability audit trails

Session 2

Fee: Rs 22,100 Apply Now

Versioned Data & Experiment Tracking

Dataset versioning & storage

DVC Pachyderm S3/GS/Blob

Experiment tracking

MLflow Weights & Biases metrics/artifacts

Ties to notebooks & code

git commits/tags env snapshots SBOM/locks

Session 3

Fee: Rs 28,000 Apply Now

Pipeline Run Metadata & Integrity

Run logs & metadata capture

Snakemake metadata Nextflow/DSL2 & Tower Galaxy histories

Integrity & trust

checksums/signatures attestations (SLSA) tamper-evident logs

Catalogs & discoverability

data catalogs schema registry lineage UI

Session 4

Fee: Rs 34,800 Apply Now

Mini Capstone: FAIR, RO-Crate & Report

Package a reproducible study artifact (RO-Crate)

Theory + Practical

Identifiers & FAIR records

DOI/ORCID **license/terms** **citation files**

Deliverables: lineage graph & audit report

graph export (JSON-LD) **PDF/HTML narrative** **manifest
& checksums**