



MESA

Multidisciplinary Environment for Scientific Advancement

IDSS PI MEETING

Goals & Strategic Plan

Project Roadmap

Tyson L. Swetnam PhD · PI (University of New Mexico)

Arthur "Barney" Maccabe PhD · David Ebert PhD · Lei Cao PhD · Co-PIs (University of Arizona)

Terrell Russell PhD Co-PI (UNC RENCi)

NSF IDSS Kick-off Category II Presentations · July 24, 2026 Virtual



THE TEAM

UNM – UA – UNC Chapel Hill



Tyson L. Swetnam, Ph.D.

Principal Investigator · UNM



Arthur B. Maccabe, Ph.D.

Co-PI · Arizona site lead



David Ebert, Ph.D.

Co-PI · Chief AI & Data Science Officer · Arizona



Lei Cao, Ph.D.

Co-PI · Lakehouse lead · Arizona



Terrell Russell, Ph.D.

Co-PI · Data Mesh Lead · RENCi Site Lead



Kory Draughn

Chief Technologist · RENCi



Illyoung Choi, Ph.D.

Research Software Engineer · Arizona



Alan King

Senior Software Developer · RENCi



Justin James

Applications Engineer · RENCi



Sarah Roberts

Research Software Engineer · Arizona



Tony Edgin

Data Curator · Arizona



Edwin Skidmore

Cloud Orchestration / Infrastructure · CyVerse



Greg Chism, Ph.D.

EOT Lead · Senior Personnel · Arizona



Jingdi Chen, Ph.D.

5G/6G Networking · Arizona



CK Chan, Ph.D.

Astronomer · Arizona



Wolfgang Jentner, Ph.D.

Human-Computer Interaction · Arizona



Junjia Guo

PhD Student · Arizona



Pengxin Wang

PhD Student · Arizona



Junyong Zhao

PhD Student · Arizona



Linan Zheng

PhD Student · Arizona

Partner Organizations & Letter Writers



University of New Mexico

CARC · lead institution · carc.unm.edu



University of Arizona

CyVerse · core site · cyverse.org



UNC-Chapel Hill

RENCI · iRODS Consortium · core site



CU Boulder

ESIL synthesis center · esiil.org



Penn State University

NCEMS synthesis center · ncems.psu.edu



Iowa State University

AIIRA — AI Institute for Resilient Agriculture



Indiana University

Jetstream-2 research cloud



TACC

Texas Advanced Computing Center



MIT CSAIL

Computer Science & AI Laboratory



Cloudflare

Industry partner



USC ISI

Information Sciences Institute · Pegasus



Event Horizon Telescope

EHT collaboration



Nokia Bell Labs

Industry partner



Carnegie Mellon

University partner



Battelle

Research & development partner



Lyrisis · DataCite

Persistent identifiers & metadata services

Multidisciplinary Environment for Scientific Advancement (MESA)

Metadata Enabled Science Agents (MESA^2!)

01

The Concept

An open source, agentic AI-driven Data Lakehouse platform that securely connects data via curated and indexed metadata.



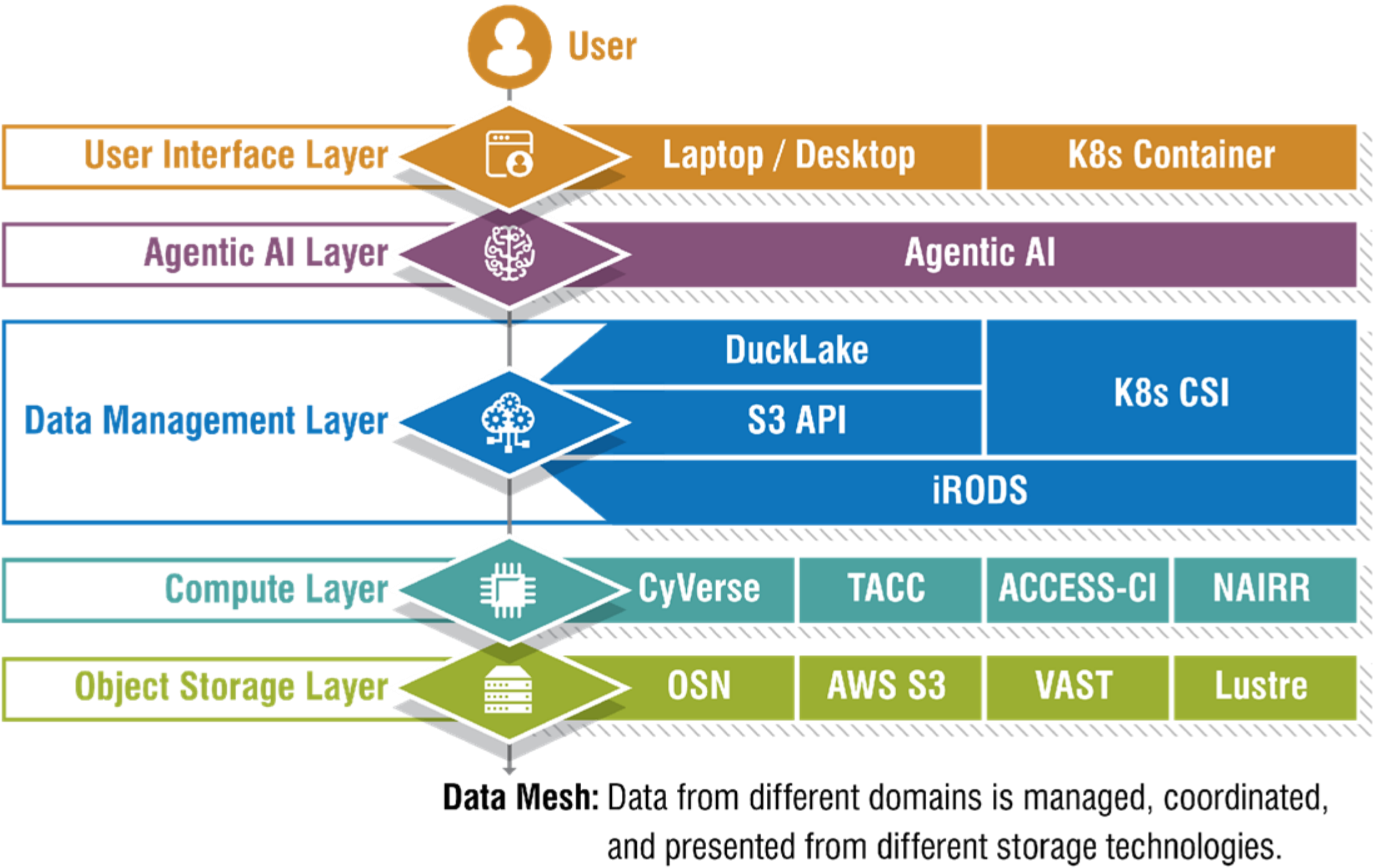
02

Platform Architecture

Five integrated layers



Five Integrated Layers



The Prompt



The Data Mesh

The Integrated Rule-Oriented Data System (iRODS) — open-source, policy-based data management from UNC's RENCI — gives MESA logical namespaces over every storage system in the mesh.

■ Data virtualization

Collections and data objects are decoupled from physical location — one namespace across OSN, AWS S3, VAST, and Lustre.

■ Data discovery

A rich, user-defined metadata catalog — the layer MESA's agents curate, index, and query to link datasets across domains.

■ Workflow automation

Rules fire at policy enforcement points — metadata extraction and curation run automatically whenever data is ingested.

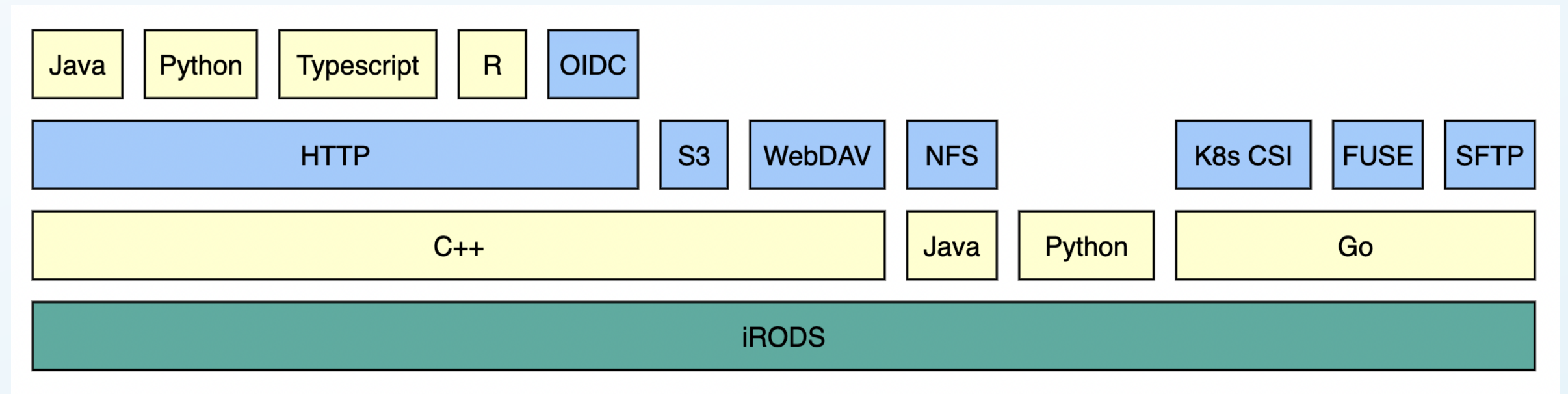
■ Secure collaboration

Tickets, permissions, and zone federation share data across institutions without moving it — how partner sites join the mesh.

One Mesh, Many Protocols

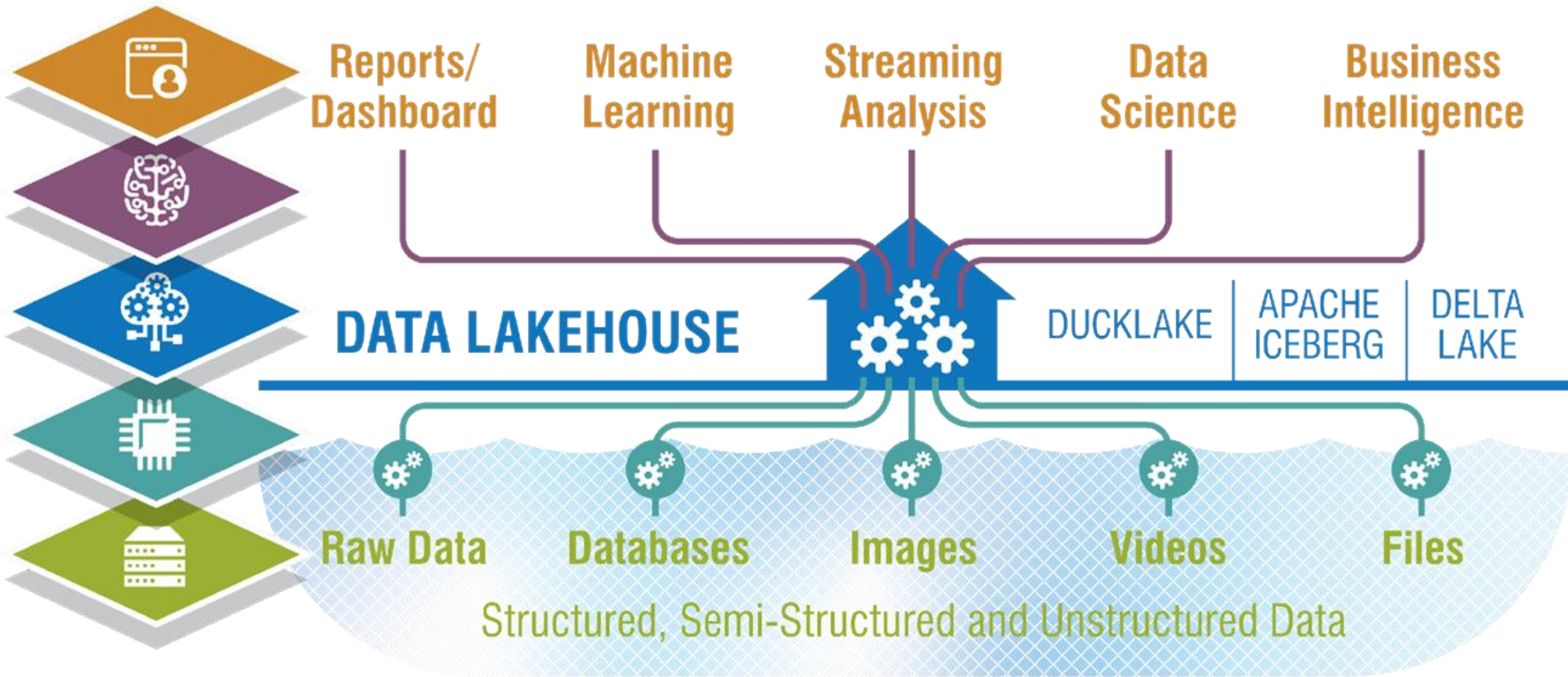
Interactions and operations handled through these layers provide the full policy enforcement, permission model, and storage abstraction as the native iRODS protocol.

All at the same time.



- DuckLake and the agentic AI layer both reach the same data through the S3 API and K8s CSI abstractions
- No copies, no format lock-in ... data stays in place while every layer of MESA speaks its preferred protocol

Lakehouse Driven Agentic Workflows



- Agents carry MCP context, RAG/OKF retrieval by request
- AI Orchestrators mediate between the Data Mesh and the Lakehouse
- Reasoning, enclaves, sandboxing keep execution secure / auditable

LIVE DEMO

MESA in Action



Demo video

■ CyVerse K8s

■ Agentic AI Sandbox + MCP tools

■ DuckLake metadata management

🌸 Clocking in for the evening shift.

How can I help you today?



Chat

Cowork

Sonnet 5 Extra ▾



Classic View



Login



NCEMS Portal

National Synthesis Center for Emergence in the Molecular and Cellular Sciences

Access cutting-edge environmental data science tools, collaborate with researchers worldwide, and accelerate your research with CyVerse computational infrastructure.

➔ Sign in with CyVerse

👤+ Create Account

📁 Browse public datasets without signing in: [View Community Data](#)

03

Roadmap



Immediate Plans — Next 6 Months

- Stand up the MESA prototype — UI & Agentic AI on CyVerse Sandboxes and Jetstream-2 VMs
- Pilot the Data Mesh — new iRODS zones at UNM and UNC RENCI; DuckLake over Open Storage Network (OSN) S3
- Release open-source MCP RAG/OKF agent framework
- Onboard anchor partners — ESIIL and NCEMS (DBI Synthesis Centers)
- Staffing core engineering team across UNM CARC, UA AI2S, and UNC RENCI

BY THE NEXT PI MEETING

Working prototype: agents answering questions against federated data in the data mesh.

Strategic Plan — 24 months



Prototype

Working platform across the three core sites



Production pilot

Anchor partners on the mesh: ESIL, NCEMS, AIIRA, EHT



Scale out

National infrastructure: NEON, NAIRR, ACCESS-CI, TACC, OSN



Cross-domain discovery

Agents linking datasets across all discovery domains



Sustainability

Open-source community governance and follow-on funding

04

Collaboration



COLLABORATION

Let's Connect!

- Contribute pilot datasets and metadata schemas from your project's domain
- Federate existing storage into the mesh — an iRODS zone or S3 bucket is enough to start
- Define a cross-project discovery Use Cases we can demonstrate together
- Use it! Give feedback ASAP

THE ASK

Pilot datasets and Use Cases



Thank You

Questions & Discussion

Tyson Swetnam, Project PI · UNM Center for Advanced Research Computing (CARC) · tswetnam@unm.edu
idss-mesa.github.io

