DATA WILL FOREVER UNDERPIN EVERY HUMAN ENDEAVOR

Mediaflux Data Platform

For Anyone With Data

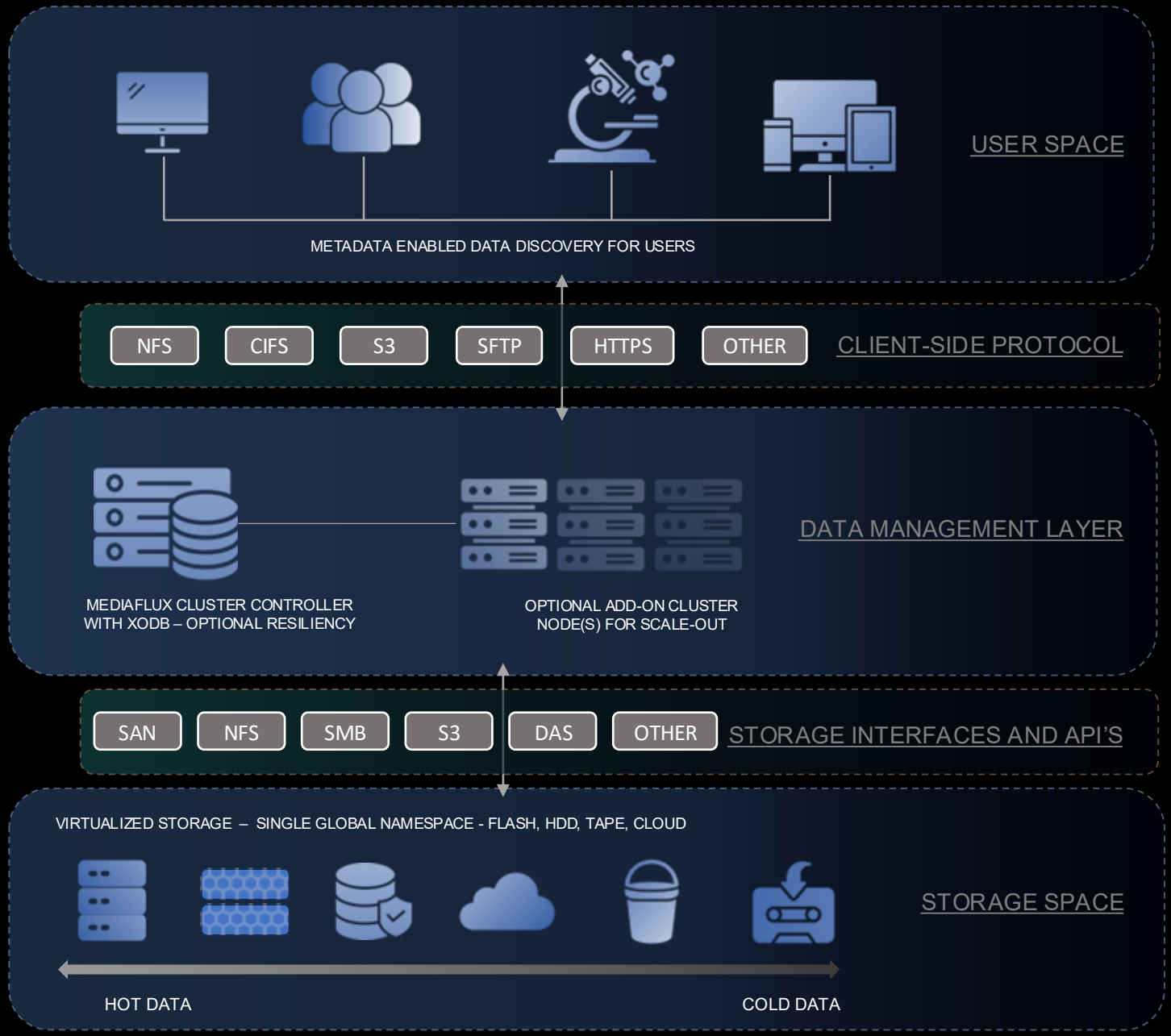
ARCITECTA

What is Mediaflux?

ARCITECTA

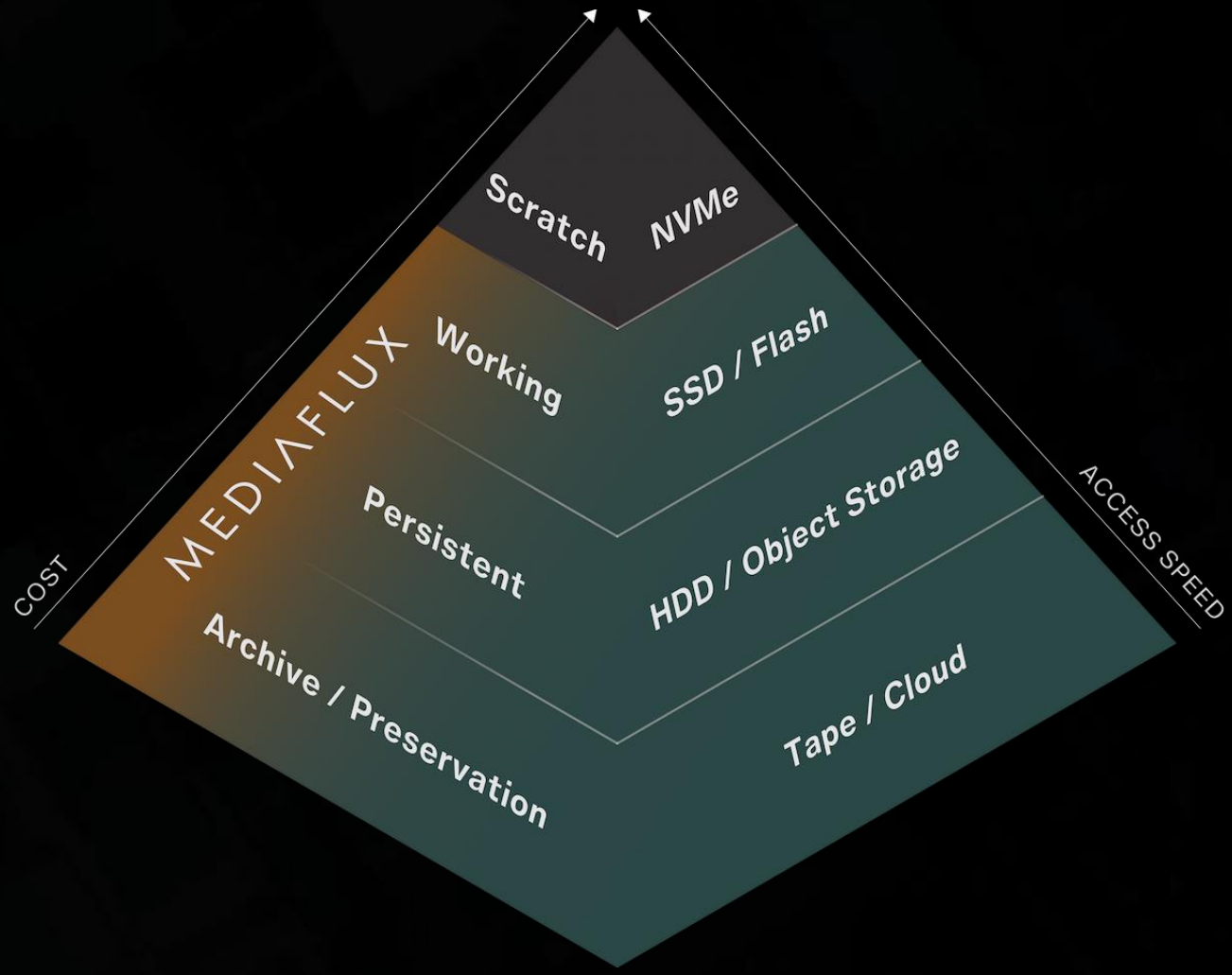
Where Mediaflux fits

- Data Rosetta stone
- Multi-protocol support enables data to be accessed by various applications
- Intelligent data placement and movement (tiering and migration)
- Extensive metadata harvesting, annotation, and cataloguing
- New Protocol connectors to today's and tomorrow's storage



Tiered Storage Defined

- Hot tier: HPC scratch, fast parallel file systems
- Warm tier: Project stores, object storage
- Cold tier: Tape, deep cloud archive
- Automate based on policy, not user memory




Mediaflux manages
Data via metadata
ARCITECTA
ARCITECTA

Types of Metadata

“System Metadata”

- File name, size, create, access, modify time, ownership permissions, etc.
- In Unix/linux world this information comes from inodes and is often used by Backup or HSM software.

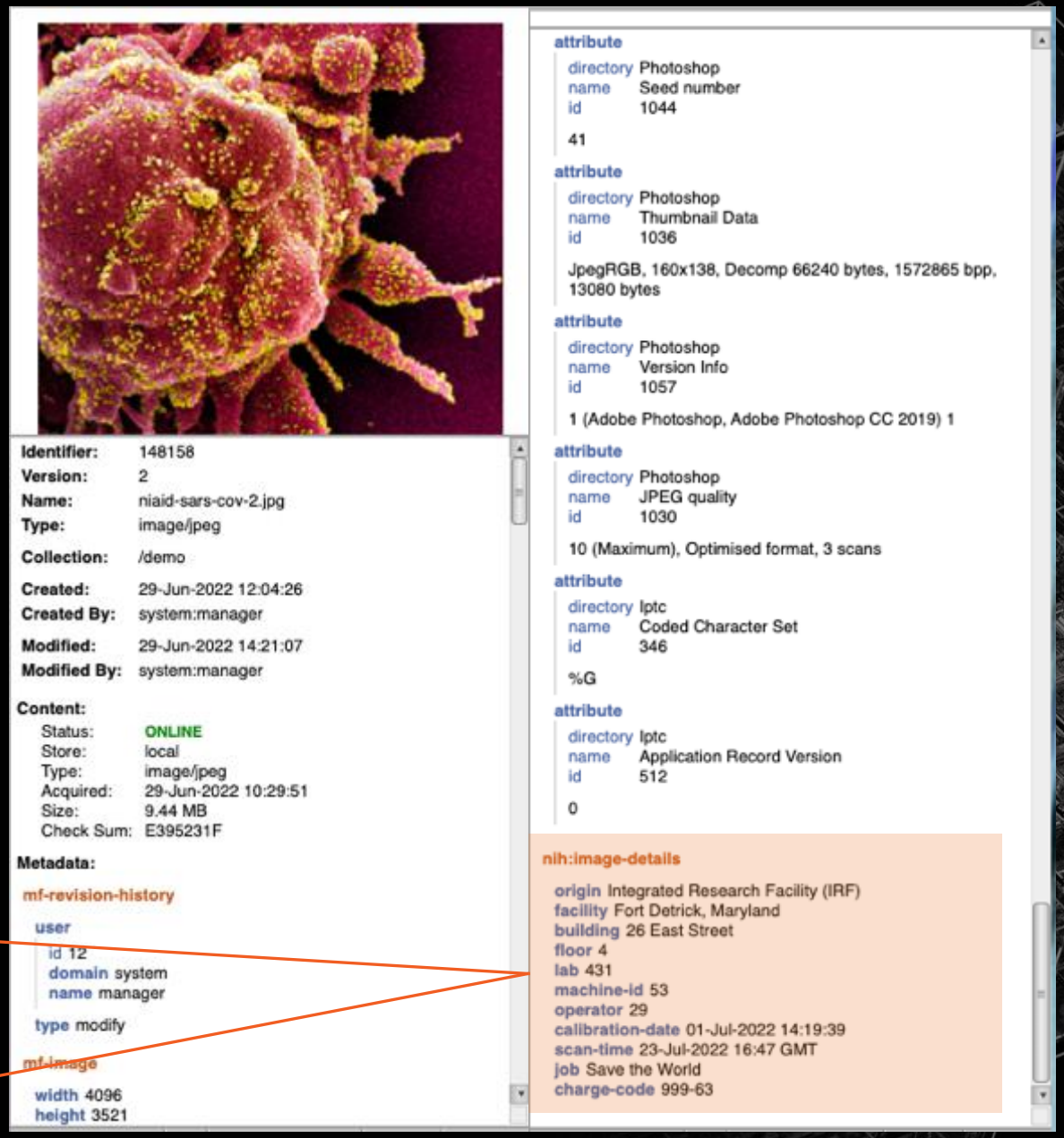
Embedded File Metadata

- Typically parsed out via MIME type.

User Defined Metadata

- This enables data life cycle management, notes, accounting

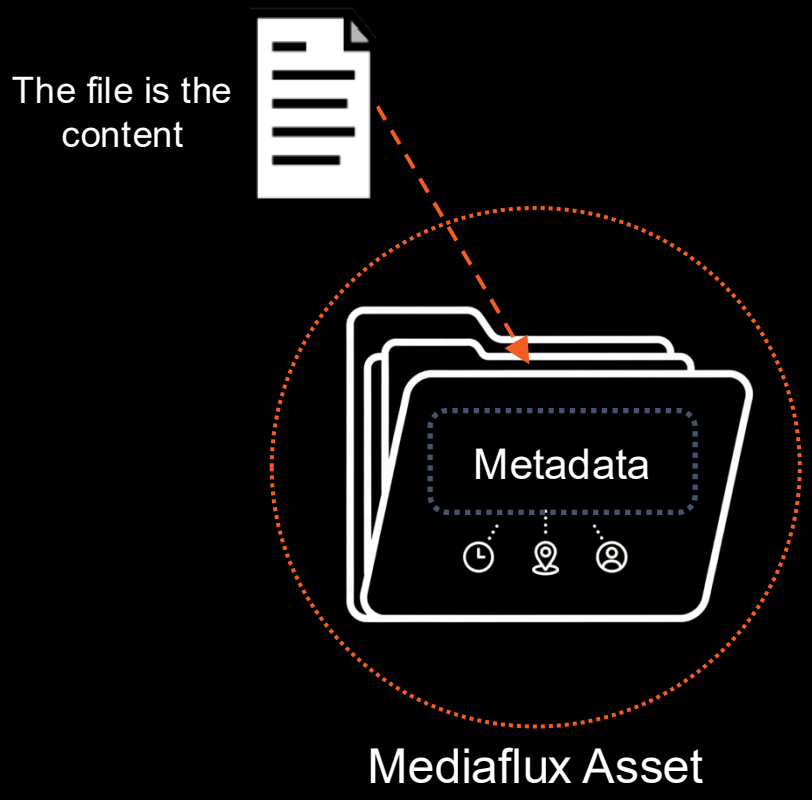
Privacy information goes here.
These fields can evolve with versioning.
Information that influences Actor/Role access models can also go here.

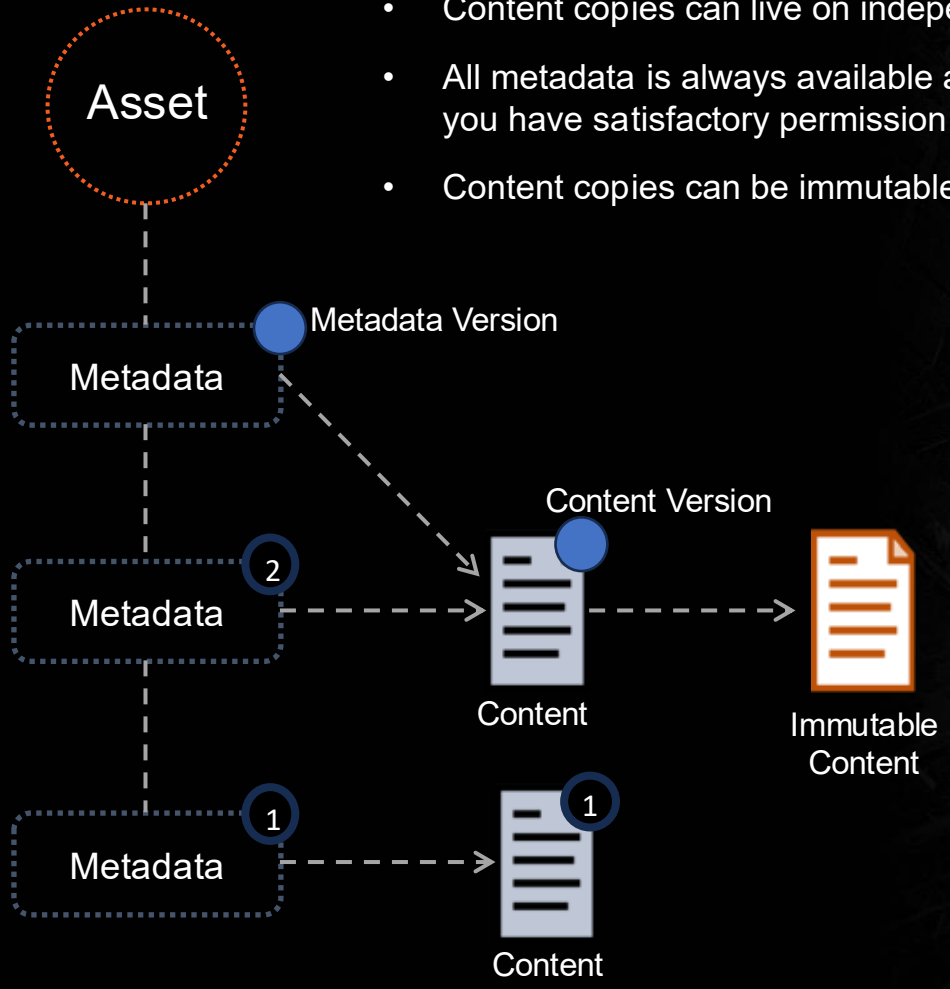


Asset Management Platform

Key Points

- The Asset is a database entity which is used to manage content
- Content = Your file
- Metadata and content are versioned independently
- Content copies can live on independent storage systems
- All metadata is always available and searchable provided you have satisfactory permission (ACL's).
- Content copies can be immutable for archives



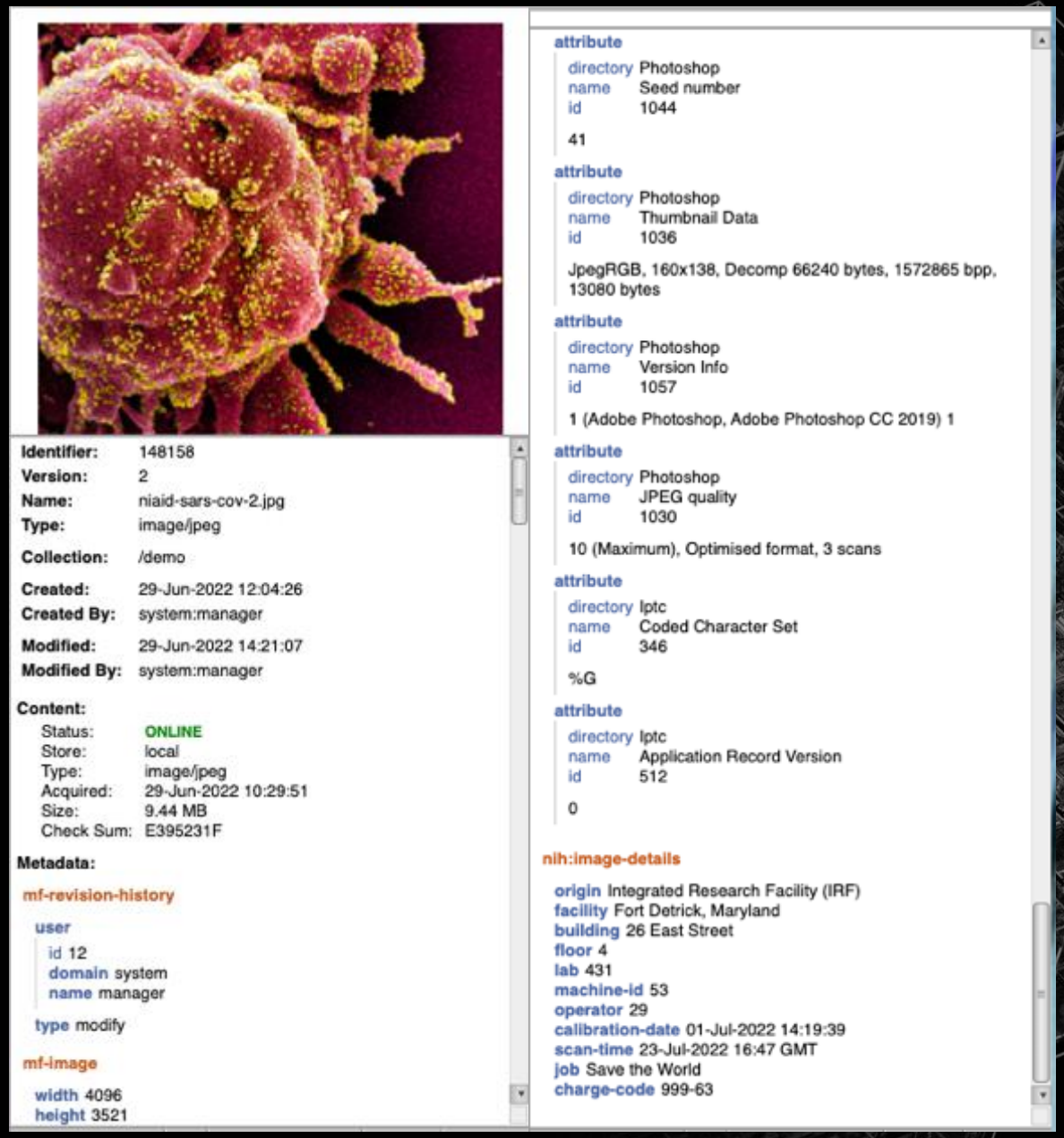


“Metadata is a love note to the future”

…Jason Scott NYPL 2011

We don’t need the perfect “form” (metadata schema).
It can be modified later, as we have a non-SQL database.

…so Write Some Metadata!



PRINCETON UNIVERSITY

100 Year Data Management Plan Challenges

Princeton Faculty, Alumni, Researchers affiliated with over 80 Nobel prize laureates

How will a Nobel prize winning researcher find their 2026 weather data sets in 2050?

Think about how many technology refreshes there will be over a 100-year period! How will they migrate all that data?

There is a small team managing 170PB of data now, the idea is to give this team “Data Power tools” to manage orders of magnitude more data over time.

Key points: always have open standard backups of content data and metadata available. No proprietary dependencies. Data should not be held hostage!

Required Metadata Updates

Roles
- Data Sponsor
- Data Manager

Project Description
- Affiliated Department(s)
- Project ID (revised)
- Project Directory
- Title
- Description

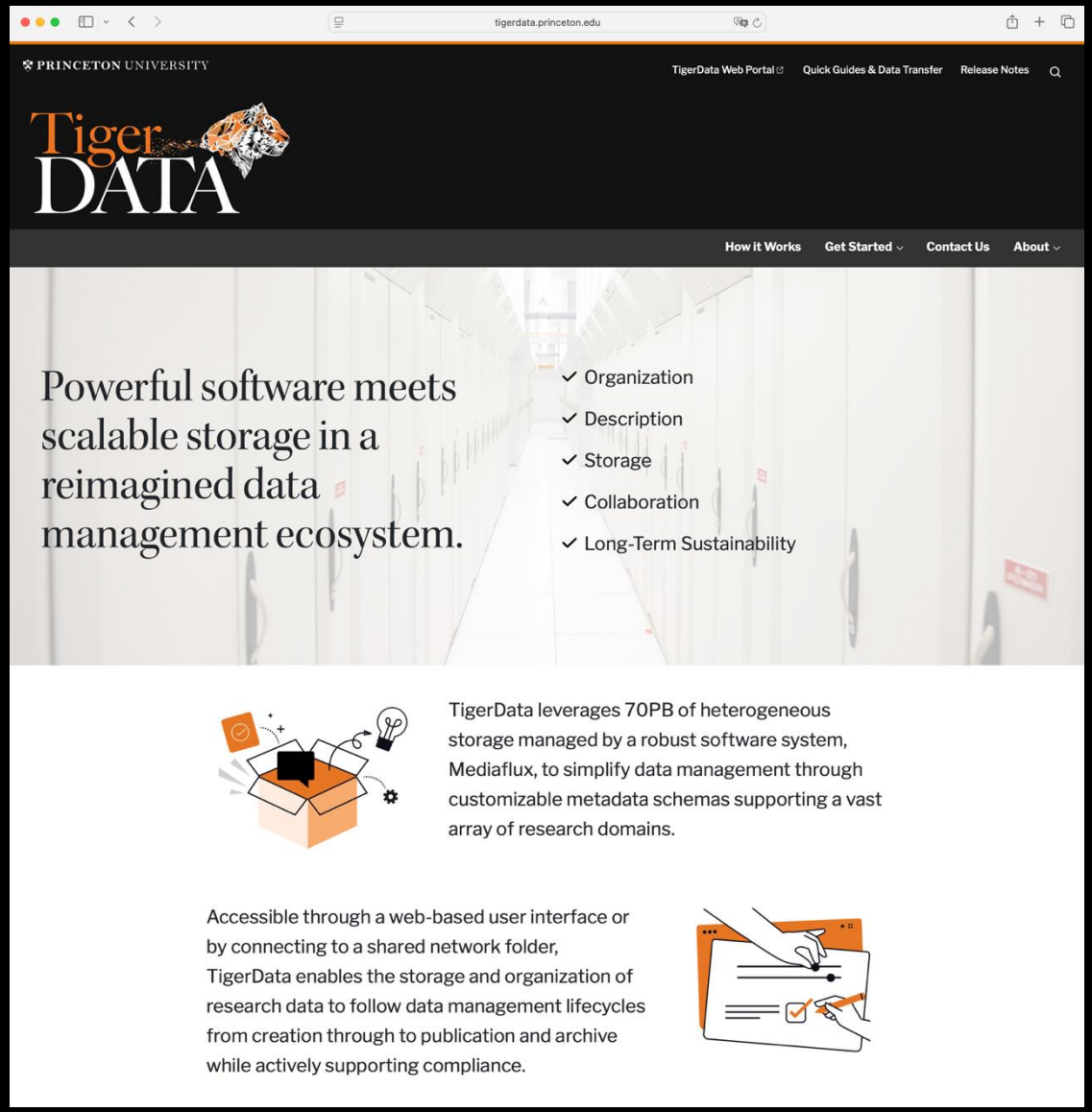
Storage and Access Needs
- Storage Capacity
- Project Visibility
- Storage Performance Expectations
- File Number Estimate
- High Performance Computing

Additional Project Information
- Project Purpose
- Provisional Project

Provenance (new)
- Submission
- Revision
- Retirement
- Publication
- Status

Example metadata from Princeton University TigerData

Resources see: Tigerdata.Princeton.edu



https://researchcomputing.princeton.edu/news/2025/introducing-tigerdata-comprehensive-data-management-service-princeton-research-community

Data Production & Analysis Is an Investment With Enduring Value

What will the future bring?

- Petabyte bricks?
- Another round of optical?
- DNA?
- Ceramic Data Storage?

Our Approach:

- New connectors to storage and users:
  - New protocols and standards (preferably open)

- I learned how to code just over 50 years ago in the age of punched cards and paper-tape
- Mediaflux was started over 25 years ago before there was S3 or Glacier
- Working Groups for new storage forming: E.g. LAST (Long-term Accessible Storage Technologies)
- Working with Cerabyte on a PILOT

Please see us at a break to see a demo video of how this could work or visit us at www.arcitecta.com

ARCITECTA®