



# USC Digital Repository

**A division of the USC Libraries**

**Created to digitally preserve academic collections at USC**

*Sam Gustman*

*Associate Dean and Chief Technology Officer*

*USC Libraries and USC Shoah Foundation*

[sgustman@usc.edu](mailto:sgustman@usc.edu)

# Overview



## Interdepartmental partnership

- USC Libraries, USC Shoah Foundation, and USC ITS.
- Over 60 Petabytes of data currently managed

## Audit services provided

- ISO 27001, SOCS 2+, MPAA, USC internal audit, Grant audits

## Highlight contracts

- Currently used by researchers in 12 departments at USC
- 250 Petabyte contract to help preserve AV collections of the Army, Navy, Airforce, and Marines
- Digitally preserve the TV and film collections of the Warner Bros. studios

# Services



**Digital preservation** – Storagetek (data center) , Filecoin (web3), Azure (cloud)

**Digitization** – Film, Video, Audio, Manuscripts of all shapes and sizes

**Cataloging and metadata**

**Digital asset management** - hosted commercial and USC DAMs

**File-server services for large assets** – 12 PB connected to USC HPC

**Interactive Interviews** – Chatbot video services

**Data Management and Repository Services** – NIH,NSF,NEH

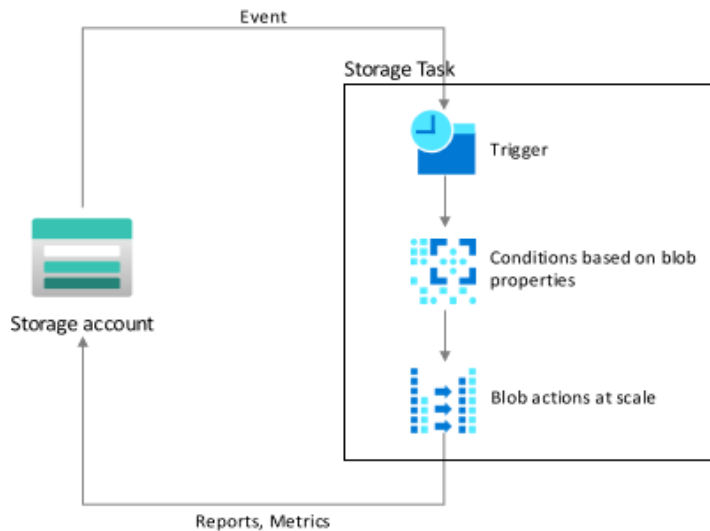
# Digital Preservation



- Four mirrored Oracle SL8500 robotic libraries around the world
- 160 Petabyte capacity
- SHA based fixity for all files
- Automatic media retirement based on age
- Mirroring of all content
- Integration with Microsoft Azure, AWS and Filecoin
- Integration with High Performance Computing center
- Integration with high speed Isilon systems for high-speed terabyte processing
- \$800 one time fee per terabyte for 20 years or \$400 one time fee per terabyte over 6 years (one tape library and one additional copy, plus more available on request).
- NSF and NIH data management plan support.



# Azure Storage Tasks (not released yet)



## Simple, no-code composition

- Framework to easily execute operations on part or all of your objects in Azure Storage
- Cost-effective execution that scales with your storage

## Capabilities

- Move objects between offline and online storage
- Manage data lifecycle and data protection
- **Run integrity checks**

## Serverless

- Fully managed infrastructure
- Trigger on schedule or on events



**Questions?**



# Additional Appendix Slides



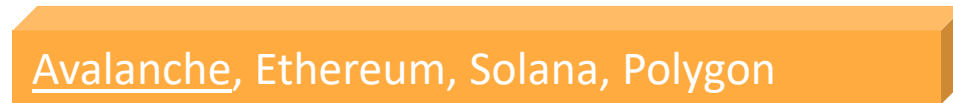
# USC Academic Filecoin Node with Web3-enabled Architecture for capture, validation, preservation and access of academic collections using blockchain technology

## Frontend



## Logic

Standard application logic  
or smart contracts

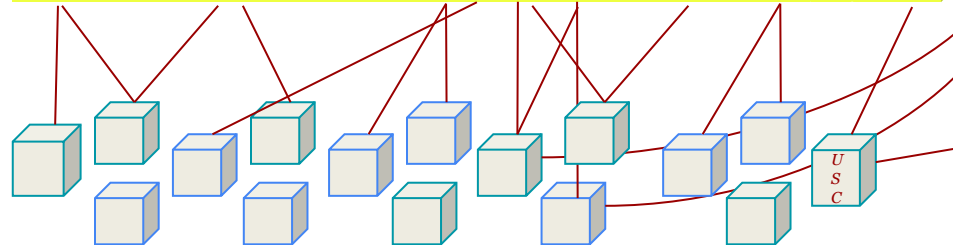


## Storage Helpers

Simplify storing & retrieving  
data (SDKs, hosted APIs, and  
more)



## Filecoin & IPFS nodes





# Cataloging and Metadata



- Indexing, cataloging, and metadata creation to enable efficient searchability and findability of digital collections
- Cost-effective cataloging support, based on the collection size, level of detail required, and availability of existing records
- Standard thesaurus or classification system, or customized vocabularies for special collections

# Security Overview



- Integrated with USC CISO including a dedicated ISO
- Audited by central USC auditing
- ISO 27001 compliant
- MPAA certified
- Fortinet 3700D Firewalls for Advanced Threat Management
- SOC2+ auditing available
- Heuristic Traffic Monitoring and Reporting
- Real-time Hardware-Based Threat Detection

# USC Clients

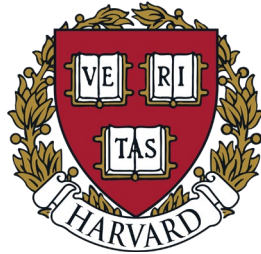


- USC Center for Applied Molecular Medicine (CAMM)
- USC Information Sciences Institute (ISI)
- USC Keck School of Medicine
- USC Marshall School of Business
- USC Thornton School of Music
- USC Rossier School of Education
- Management Information Systems
- USC Ming Hsieh Department of Electrical Engineering - Signal and Imaging Processing Institute (SIPI)
- USC UPC Genome Core
- USC ITS
- USC Center for Religion and Civic Culture
- USC Department of Psychology - Stephen Read
- USC Libraries
- USC School of Dentistry - Stephen Yen
- USC Dornsife - Minds in Development Lab
- USC Thornton School of Music

# External Clients



Academy of Motion  
Pictures and Sciences  
(Oscar Awards)



Harvard Judaica



GLAAD Media Awards



Television Academy  
(Emmy Awards)



Go For Broke National  
Education Center



United States  
Department of  
Defense



Wende Museum's  
Albanian Human  
Rights Collection



Warner Bros.  
Entertainment Inc.  
(WB) (WB)

# Fulfilling Data Management Plan (DMP) Requirements



The USC Digital Repository offers a full suite of data management services to help researchers meet new requirements for data preservation, access, and security by the National Institutes of Health (NIH), the National Science Foundation (NSF), and other federal agencies.

<https://repository.usc.edu/nih-data-management-and-sharing-repository-services>

Our data management services and the technology supporting those services are:

- Auditable
- Customizable
- Undergoing SOC2 Compliance Certification (readiness assessment completed)

NIH proposal submitted by Dr. Steven Lin with USCDR DMP that includes:

- Collection/Data Assessment and Review
- Digital Library Account and Access Setup for each additional 300 files
- Metadata/Cataloging
- Access Review and Request Fulfillment



# Interactive Interviews

Based on the USC Shoah Foundation's Dimensions in Testimony project, the USC Digital Repository now offers service to anyone of interest to the USC community to create their own interactive interviews. The first pilot was Alice and the Cheshire Cat released last year. The first live interview is of Lora King, Rodney King's daughter, for Annenberg's USC Charlotta Bass Journalism & Justice Lab.

