

PROJECT MYCELIUM: NEXT GENERATION DISTRIBUTED DIGITAL PRESERVATION

Nathan Tallman

Penn State University

ntt7@psu.edu

@ntallman@digipres.club

Glenn Heinle

KeeperTechnology

glenn.heinle@keepertech.com

Introduction



- MetaArchive started as a project in 2004 and became a distributed digital preservation service provider in 2007.
- Since the beginning, MetaArchive has used Stanford's Lots of Copies Keep Stuff Safe (LOCKSS) to provide bit-level preservation.
- LOCKSS is great for tightly controlled networks and static journal content. But MetaArchive embraces administrative diversity with distributed network management and members deposit a wide variety of cultural heritage materials, in a variety of packaging standards.
- LOCKSS has served us well but is now inhibiting network growth.

Triple Bottom Line Sustainability and UN Sustainable Development Goals

- Triple Bottom Line Sustainability
 - Environmental Sustainability (Planet)
 - Economic Sustainability (Prosperity)
 - Labor Sustainability (People)
- UN Sustainable Development Goals
 - Goal 9: Reduced inequality within and among countries



9 INDUSTRY, INNOVATION
AND INFRASTRUCTURE



**BUILD RESILIENT INFRASTRUCTURE, PROMOTE INCLUSIVE AND
SUSTAINABLE INDUSTRIALIZATION AND FOSTER INNOVATION**

Research and Development

- Partnership with Keeper Technology
- Multi-phased Approach
 - 1: Requirements gathering, system design
 - 2: Prototyping, proof-of-concept
 - 3: Deployment of a pilot network



Discovery

Organization	Department	Member Name	Member Title	Member Email	Member Phone	Member Location	Member Timezone	Member Availability	Member Skills	Member Interests	Member Bio	Member Photo	Member Signature	Member Avatar	Member Profile	Member Bio	Member Photo	Member Signature	Member Avatar	Member Profile
University of Kentucky	Computer Science	John Doe	Professor	john.doe@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	John Doe is a professor at the University of Kentucky, where he has been teaching for over 20 years. He has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									
University of Kentucky	Computer Science	Jane Smith	Assistant Professor	jane.smith@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	Jane Smith is an assistant professor at the University of Kentucky, where she has been teaching for over 10 years. She has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									
University of Kentucky	Computer Science	Bob Johnson	Associate Professor	bob.johnson@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	Bob Johnson is an associate professor at the University of Kentucky, where he has been teaching for over 15 years. He has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									
University of Kentucky	Computer Science	Alice Brown	Assistant Professor	alice.brown@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	Alice Brown is an assistant professor at the University of Kentucky, where she has been teaching for over 10 years. She has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									
University of Kentucky	Computer Science	Charlie White	Associate Professor	charlie.white@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	Charlie White is an associate professor at the University of Kentucky, where he has been teaching for over 15 years. He has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									
University of Kentucky	Computer Science	Diana Green	Assistant Professor	diana.green@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	Diana Green is an assistant professor at the University of Kentucky, where she has been teaching for over 10 years. She has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									
University of Kentucky	Computer Science	Frank Black	Associate Professor	frank.black@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	Frank Black is an associate professor at the University of Kentucky, where he has been teaching for over 15 years. He has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									
University of Kentucky	Computer Science	Grace King	Assistant Professor	grace.king@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	Grace King is an assistant professor at the University of Kentucky, where she has been teaching for over 10 years. She has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									
University of Kentucky	Computer Science	Henry Lee	Associate Professor	henry.lee@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	Henry Lee is an associate professor at the University of Kentucky, where he has been teaching for over 15 years. He has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									
University of Kentucky	Computer Science	Ivy Hill	Assistant Professor	ivy.hill@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	Ivy Hill is an assistant professor at the University of Kentucky, where she has been teaching for over 10 years. She has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									
University of Kentucky	Computer Science	Jack Stone	Associate Professor	jack.stone@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	Jack Stone is an associate professor at the University of Kentucky, where he has been teaching for over 15 years. He has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									
University of Kentucky	Computer Science	Karen Wood	Assistant Professor	karen.wood@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	Karen Wood is an assistant professor at the University of Kentucky, where she has been teaching for over 10 years. She has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									
University of Kentucky	Computer Science	Leo Hill	Associate Professor	leo.hill@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	Leo Hill is an associate professor at the University of Kentucky, where he has been teaching for over 15 years. He has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									
University of Kentucky	Computer Science	Mia Hill	Assistant Professor	mia.hill@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	Mia Hill is an assistant professor at the University of Kentucky, where she has been teaching for over 10 years. She has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									
University of Kentucky	Computer Science	Noah Hill	Associate Professor	noah.hill@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	Noah Hill is an associate professor at the University of Kentucky, where he has been teaching for over 15 years. He has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									
University of Kentucky	Computer Science	Olivia Hill	Assistant Professor	olivia.hill@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	Olivia Hill is an assistant professor at the University of Kentucky, where she has been teaching for over 10 years. She has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									
University of Kentucky	Computer Science	Peter Hill	Associate Professor	peter.hill@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	Peter Hill is an associate professor at the University of Kentucky, where he has been teaching for over 15 years. He has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									
University of Kentucky	Computer Science	Quinn Hill	Assistant Professor	quinn.hill@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	Quinn Hill is an assistant professor at the University of Kentucky, where she has been teaching for over 10 years. She has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									
University of Kentucky	Computer Science	Rachel Hill	Associate Professor	rachel.hill@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	Rachel Hill is an associate professor at the University of Kentucky, where she has been teaching for over 10 years. She has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									
University of Kentucky	Computer Science	Sam Hill	Associate Professor	sam.hill@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	Sam Hill is an associate professor at the University of Kentucky, where he has been teaching for over 15 years. He has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									
University of Kentucky	Computer Science	Tina Hill	Assistant Professor	tina.hill@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	Tina Hill is an assistant professor at the University of Kentucky, where she has been teaching for over 10 years. She has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									
University of Kentucky	Computer Science	Uma Hill	Associate Professor	uma.hill@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	Uma Hill is an associate professor at the University of Kentucky, where she has been teaching for over 10 years. She has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									
University of Kentucky	Computer Science	Victor Hill	Associate Professor	victor.hill@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	Victor Hill is an associate professor at the University of Kentucky, where he has been teaching for over 15 years. He has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									
University of Kentucky	Computer Science	Wendy Hill	Assistant Professor	wendy.hill@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	Wendy Hill is an assistant professor at the University of Kentucky, where she has been teaching for over 10 years. She has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									
University of Kentucky	Computer Science	Xavier Hill	Associate Professor	xavier.hill@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	Xavier Hill is an associate professor at the University of Kentucky, where he has been teaching for over 15 years. He has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									
University of Kentucky	Computer Science	Yara Hill	Assistant Professor	yara.hill@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	Yara Hill is an assistant professor at the University of Kentucky, where she has been teaching for over 10 years. She has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									
University of Kentucky	Computer Science	Zoe Hill	Associate Professor	zoe.hill@uky.edu	606-253-2234	Lexington, KY	EST	Available	Python, Java, JavaScript	Research in AI and ML	Zoe Hill is an associate professor at the University of Kentucky, where she has been teaching for over 10 years. She has a PhD in Computer Science from the University of Michigan and has published several papers in the field of AI and ML.									

Questionnaire

Tuesday, January 13, 2025 1:29 PM

Intent

Send out the questionnaire to all MetaArchive members, then setup interactive sessions with two or more active members. Ideally, we would be able to talk to one member with a cache and one without. The big section "Questions we want..." would be forwarded to MetaArchive members. We can dig into some of these questions into the live interaction as well as probe items in the "Stuff we need to know" section.

- Work to modify the questions to include examples or better descriptions or examples on technical areas
 - How does a user get data out of the system?

Questions we want to ask of the MetaArchive members:

<https://github.com/locks/locks-diemon>

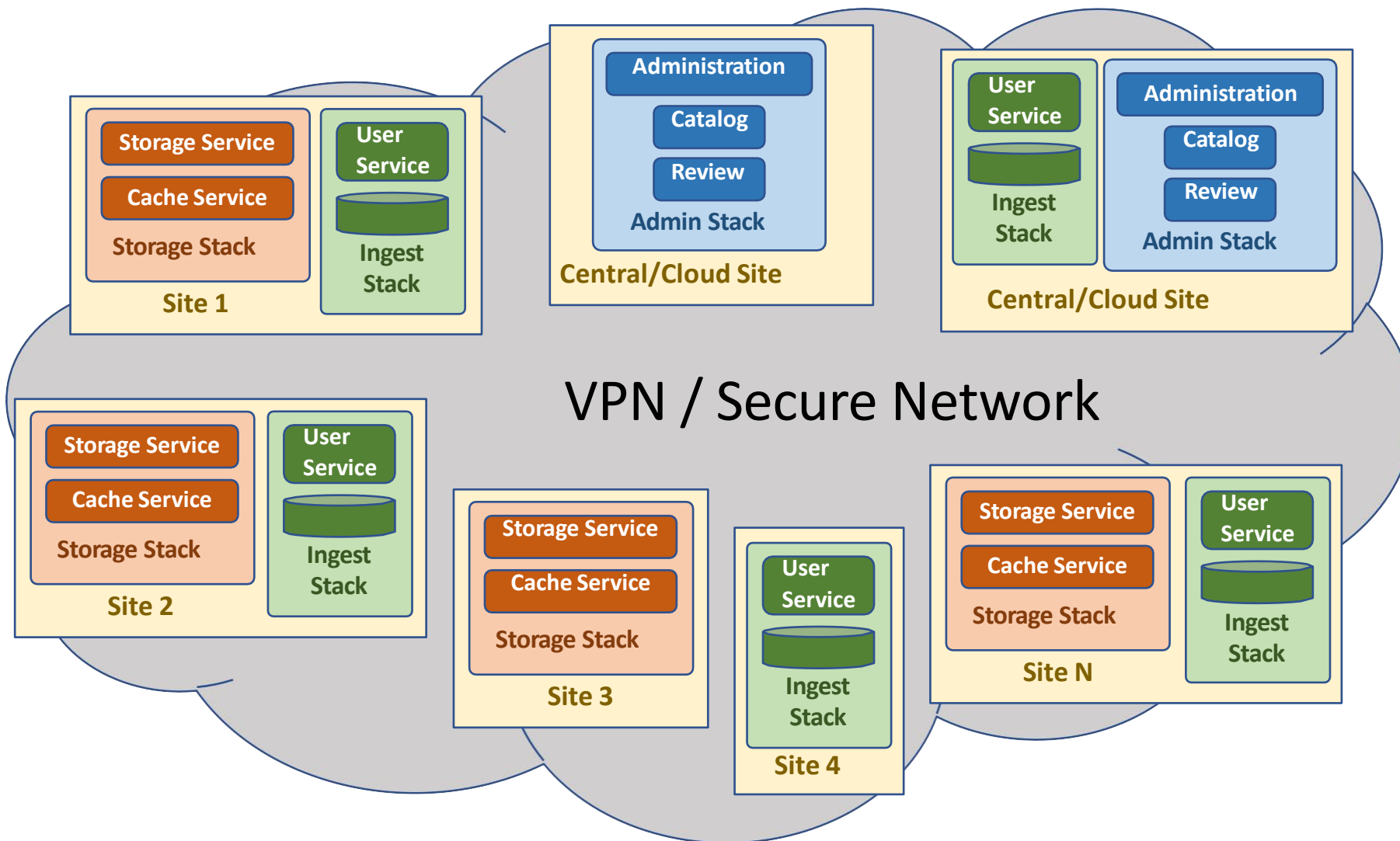
- Preferred way to have the system ingest data? (these might need more explanation or examples)
 - FTP/SFTP?
 - API?
 - Web?
- What level of granularity is desired for access and sharing? (Examples might help)
 - If that didn't exist, would you preserve more data into the system?
- What kind of (descriptive) metadata do you enter to keep track of assets?
 - What could be done to make this easier?
- How large are the archive objects you want to submit to the storage environment?
 - Min size? Max size? Average size?
 - How many such objects do you expect to submit?
 - As of April 2021, members said they expect (hope) to ingest 43-157 TB within the next three years (2022-2025)
 - We have less than 25 TB of unique data in the network now
 - Highest Annual growth rate ever is about 4/7th year
- How do you define where your data can be stored? (Which member caches?)
 - Do you want to have more control on where your data is stored?
 - How? this is not currently controlled/specified by members. Target caches are assigned by central staff based on cache capacity and geographic distribution. You could maybe ask if they would like to have this capability?
- Should the login credentials from your organization be utilized for gaining access to the preservation system?
 - Is there an asset identifier?
 - How? Does this mean, are there any assets that do not require indefinite preservation or have a set retention period?
- Please tell us a little bit about your current (HW/lock?) infrastructure.
 - How? I think that this question would be best answered by member IT staff, potentially in consultation with members who ingest content to the network, but something to warn them about when we distribute the survey (or potentially this is a set of questions that should go in a different survey/focus group)
 - Amount and type of storage (might need explanation or examples)
 - AWS, S3, etc.
 - Network connectivity to peer sites
 - User authentication
 - LDAP/S, authentication use
 - How? Depending on what you'd like to know about currently workflow, another set of tools to ask about are ones for creating Back tags - Baggar, Exactly, DART (GUTH) and bagit.py (GUT)
 - Would you be interested in participating in a follow up interview with BaggarTech to further discuss your workflow and needs?

• Responses:

- Difficult to Use
- Understaffed
- Automation
- Ease of Ingest
- Physical Shipment

- Wide variety of expertise, infrastructure, workflow

Architecture

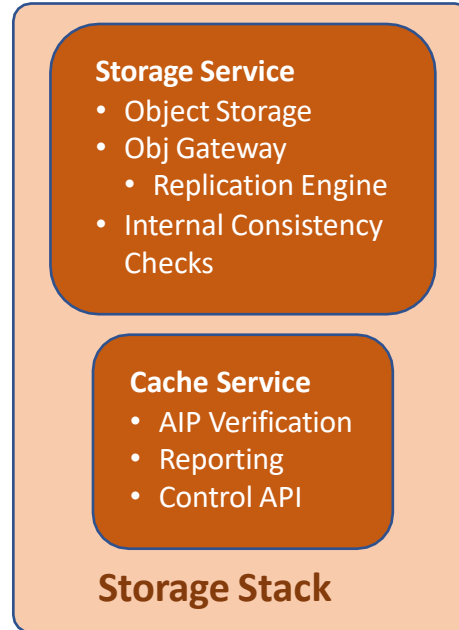
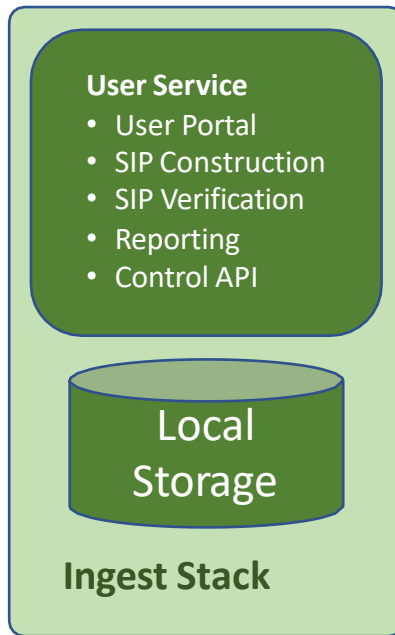


- Goals:
 - Easy
 - Automation
 - Increase Usage
- Distributed Architecture
 - Across Member Sites
- Modular Design
 - Well Defined Interfaces
- Scalable
- Secure
- Extensible

Components

Ingest Stack

- Primary User Interface
- Local or Cloud
- Archive-Ready Packages and Native Assets

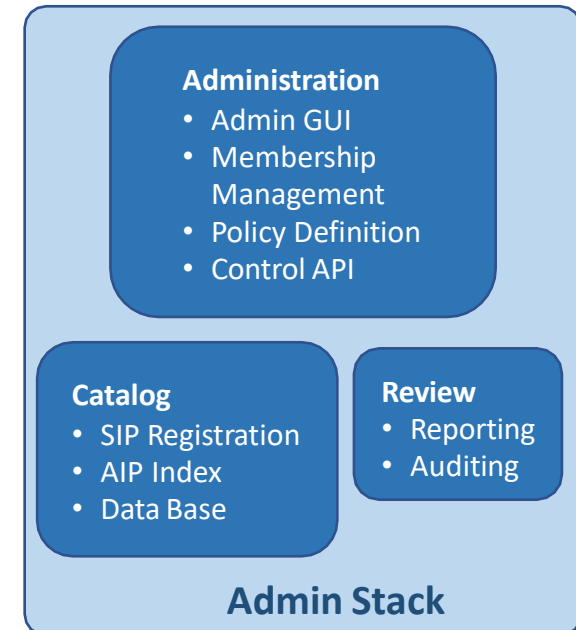


Storage Stack

- Data Protection
- Internal Consistency Checks
- External Fixity Verification
- Content Un-Aware

Admin Stack

- Limited Access
- Globally Unique Registration
- Additional Metadata
- Policies, Membership, Etc.



THANK YOU

Nathan Tallman

Penn State University

ntt7@psu.edu

@ntallman@digipres.club

Glenn Heinle

KeeperTechnology

glenn.heinle@keepertech.com