

One to Many: Connecting Local Repository Systems with Distributed Digital Preservation Systems

Sibyl Schaefer
UC, San Diego

Bill Branan
LYRASIS

Problem Statement

Local repository systems are active: DSpace, Fedora, Hyrax

Distributed Digital Preservation systems (DDPs) are fairly static and often have limited versioning capabilities:

Chronopolis, APTrust, LOCKSS

Information about the data in the DDP (location, audit) is not captured in the local repository space with the rest of the metadata.

Use Case

UC San Diego sends about 45 TB of data from its local repository to Chronopolis quarterly.

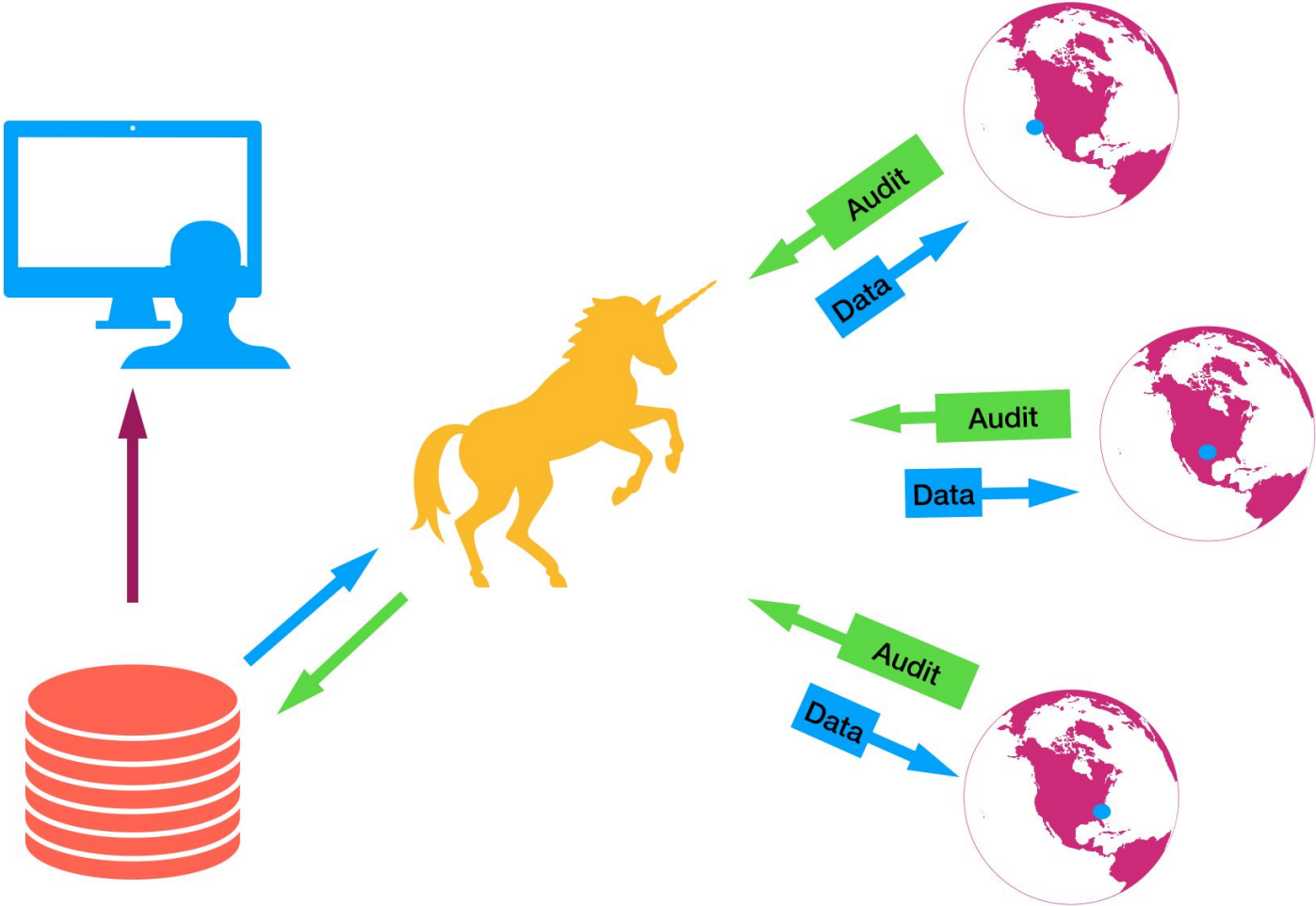
This currently takes about **30 days**, almost two thirds of which is for data packaging (tarring, bagging).

Actual changes to the entire data set tend to be minor - metadata changes or new data added. Usually only a max of 6 TB difference.

Goals

- 1) To define the development work needed to integrate local repositories and DDPs
- 2) To define requirements for an interface for curators to send digital objects from their local repository to a DDP
- 3) To define the requirements for version information and tracking of data sent to a DDP service
- 4) To ensure that the created definitions, specifications, and design documents are applicable to other digital repository software and DDP services

Magic!



Teams

Core Team

- Sibyl Schaefer (UC San Diego)
- Jessica Hilt (UC San Diego)
- Mike Ritter (University of Maryland)
- David Trujillo (UC San Diego)
- Andrew Woods (Lyrasis)
- Bill Branan (Lyrasis)
- Tom Johnson (UC Santa Barbara)
- Rosalyn Metz (Emory University)

Advisory Team

- Tim Marconi (UC San Diego)
- Erin Glass (UC San Diego)
- David Minor (UC San Diego)
- Andrew Diamond (APTrust)
- Collin Brittle (Emory University)
- Brendan Quinn (Northwestern University)
- Tom Wrobel (Oxford University)

User Stories

- Capture these features/functions:
 - Send all your data to a DDP or multiple DDPs
 - Send partial amounts of data to a DDP or multiple DDPs
 - Configure workflows for when data should be sent
 - Send versions of data
 - Send updated metadata
 - Remove particular objects or files from the DDP*
 - Retrieve/restore data from the DDP
 - Tracking audit info from the DDP - if fixity checks were OK, etc.

**according to policy*

Repository

Hyrax
Fedora
DSpace
...

Manual
Transfers



Distributed Digital Preservation (DDP) System

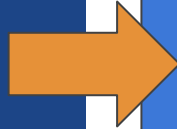


Chronopolis
APTrust
LOCKSS
...

Repository

Gateway

Hyrax
Fedora
DSpace
...

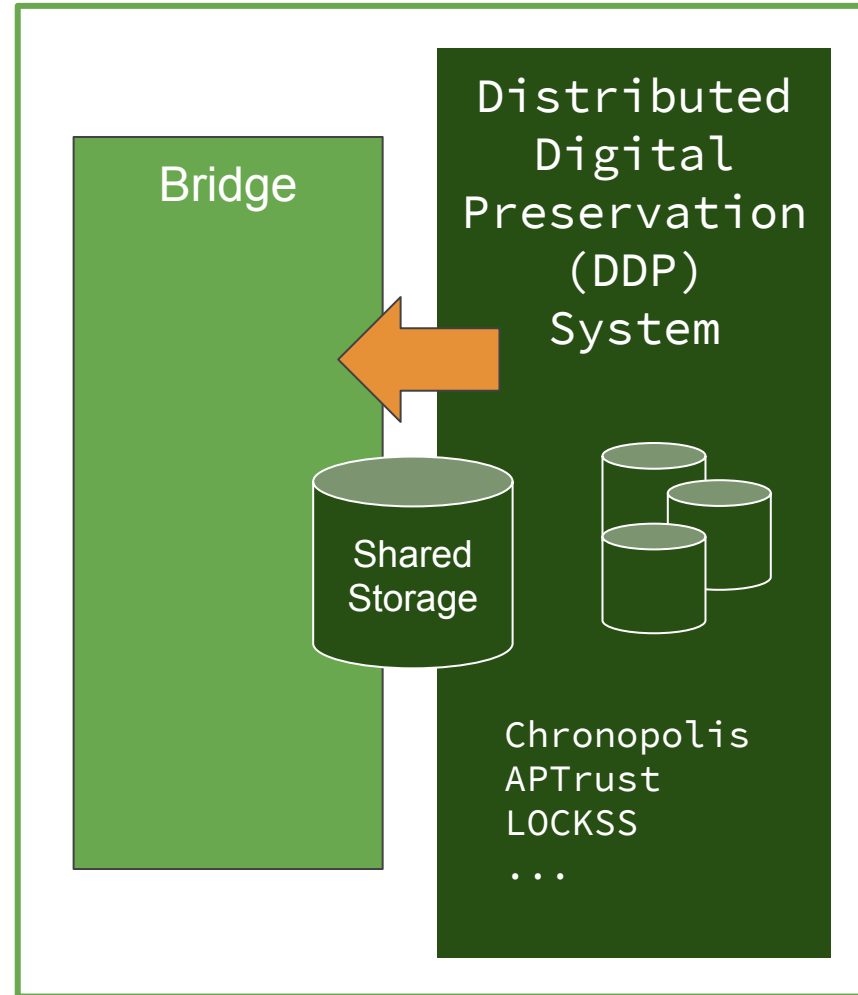


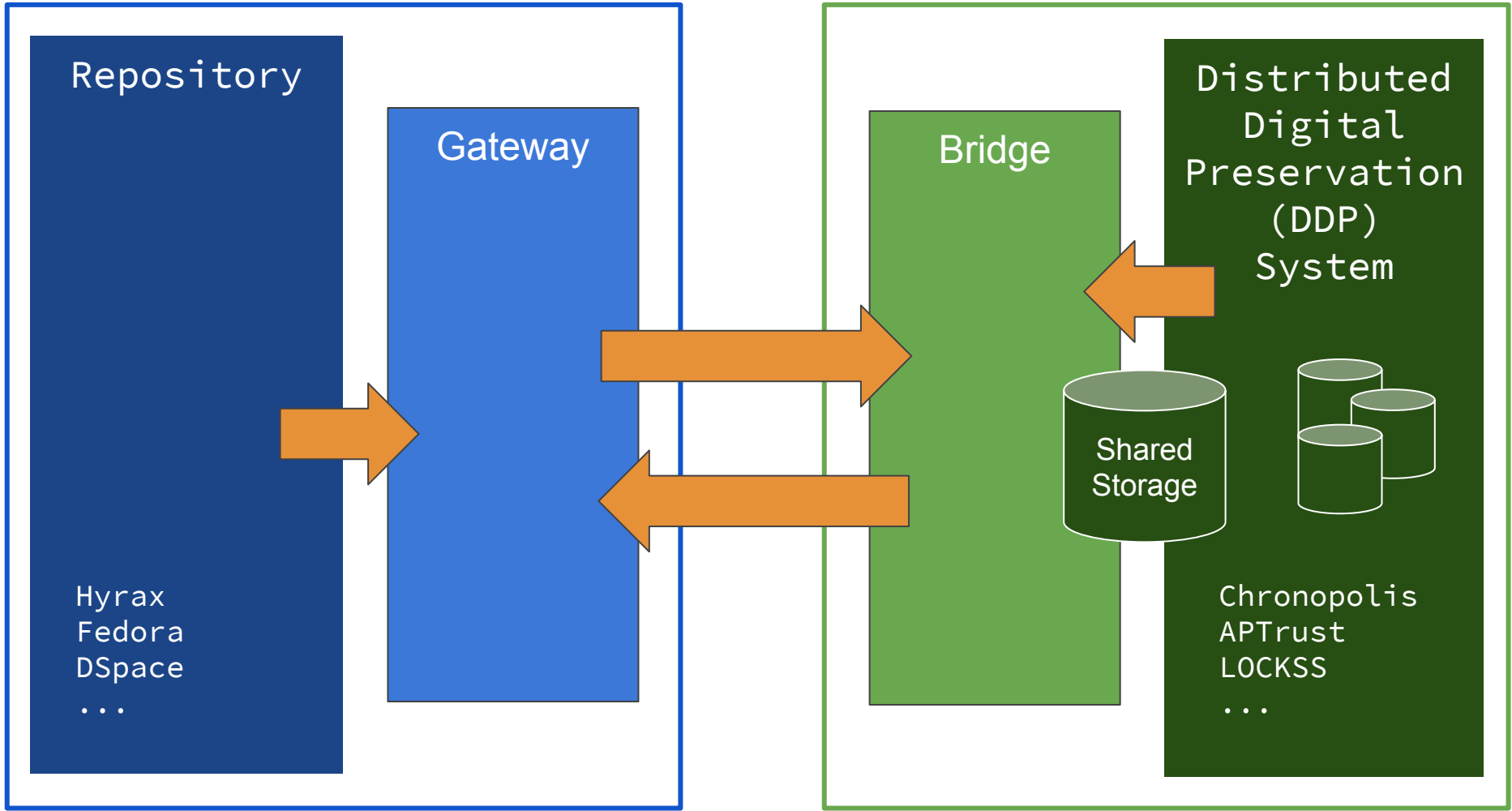
Gateway

- Pathway to preservation services
- Deployed alongside a repository
- API called by the repository
- Aggregating cache for preservation requests
- Synchronous interaction with the repository
- Able to connect to multiple preservation systems
- Version aware

Bridge

- Deployed alongside a DDP
- Provides an API to be called by the Gateway
- Provides an API to be called by the DDP
- Shares staging storage with DDP
- Able to connect to multiple systems (not only repositories)
- Provides status for deposit, restore, and delete actions
- Provides access to audit log for all deposited content
- Version aware





Benefits

- Allows automation of the deposit and restore workflows
- Allows repository to deposit into multiple DDPs
- Allows DDP to easily accept content from many depositors
- Provides direct access to details about deposited content
- Provides synchronous responses for the repository
- Allows optimized file transfer
- Supports direct restore and deletion requests
- Agnostic to file and transport packaging formats
- Supports versioning, including delta deposits

Progress and Next Steps

— — —

- Progress:
 - User stories created
 - Overall architecture determined
 - UI diagrams for Hyrax drafted
 - Specifications drafted
- Next Steps:
 - Complete Specifications
 - Finalize User Interface diagrams
 - Planning for follow-on development phase
- Grant completion in Q1 2020

More information

Project goals, user stories, and draft specifications available at: <https://wiki.duraspace.org/display/OTM>

Questions/Comments? Please contact Sibyl: sschaefer@ucsd.edu

This grant project was made possible by funding from the Andrew W. Mellon Foundation.