

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

**Sibyl Schaefer
University of California, San Diego**

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.

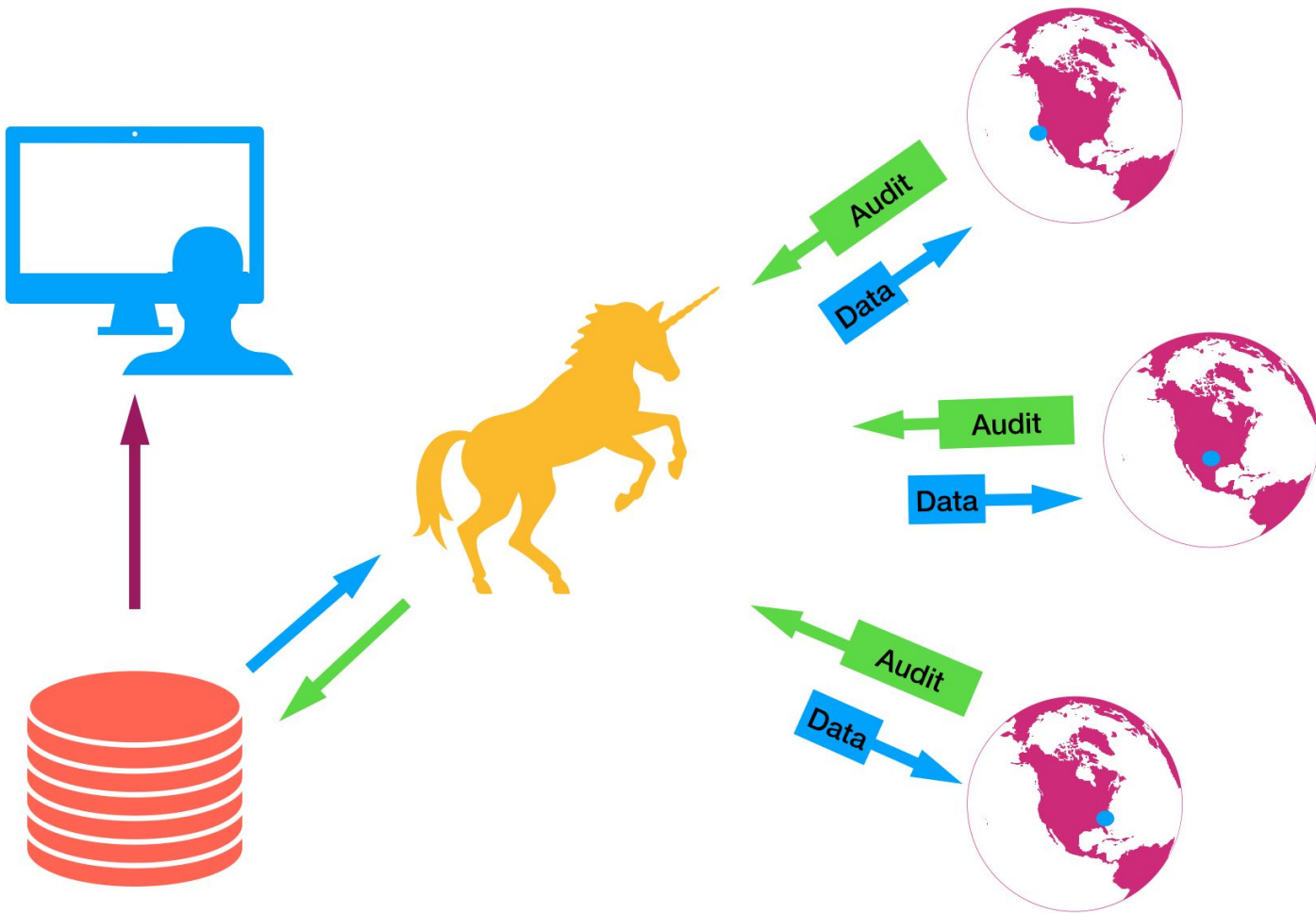
This currently takes about **30 days**, almost two thirds of which is for data packaging (tarring, bagging).

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 Brannan (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)

Progress

- User stories created including an open period where comments by the community were welcomed and then incorporated.
- Overall architecture (the unicorn) determined.
- Specifications drafted and available for comment.

Challenges

Our fourth goal - making sure that the specifications developed are applicable to other repository software and DDPs - has been the most challenging because it reduces the amount of feature complexity one can expect in either the local repository or the DDP.

More information

Project goals, use cases, and draft specifications available at:

<https://wiki.duraspace.org/display/OTM>

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