

# INTEGRATING DURACLOUD WITH DPN AT CHRONOPOLIS & THE TEXAS DIGITAL LIBRARY

---

Bill Branan

Sibyl Schaefer

Ryan Steans

Open Repositories 2015





---

Bill Branan, DuraCloud Technical Lead



# DuraCloud

## Hosted digital preservation service

- Distributed off-site storage
- Extremely large storage capacity
- Automated duplication
- Automated synchronization tooling
- Verified data integrity
- Simple UI administration
- Powerful integration options
- Comprehensive REST API
- Predictable annual billing
- Personalized support
- Open to anyone
- Open source code



# Reasons to Partner

## Shared beliefs

- Digital content preservation is important to the future of society
- Digital preservation needs to be easier to accomplish
- A digital preservation solution must be economically viable
- Need to support preservation needs of all institutions, regardless of size or technical capability

## Compatible organizational strengths

# Shared Solution



## Purpose of DPN

- Formed to ensure long term preservation of the digital scholarly record
- Protects against catastrophic loss due to technology, organization, or natural disasters
- Provides a financial model for the preservation of content over time
- For the academy, by the academy

# Structure of DPN



- 5 founding nodes in preservation federation
  - More nodes to be added over time
- Nodes differ from one another in:
  - Underlying storage technology
  - System administration techniques
  - Geographical location
- Nodes may choose to receive data from external institutions, from other DPN nodes, or both
- Each node may offer additional features and functionality

# Two DPN nodes

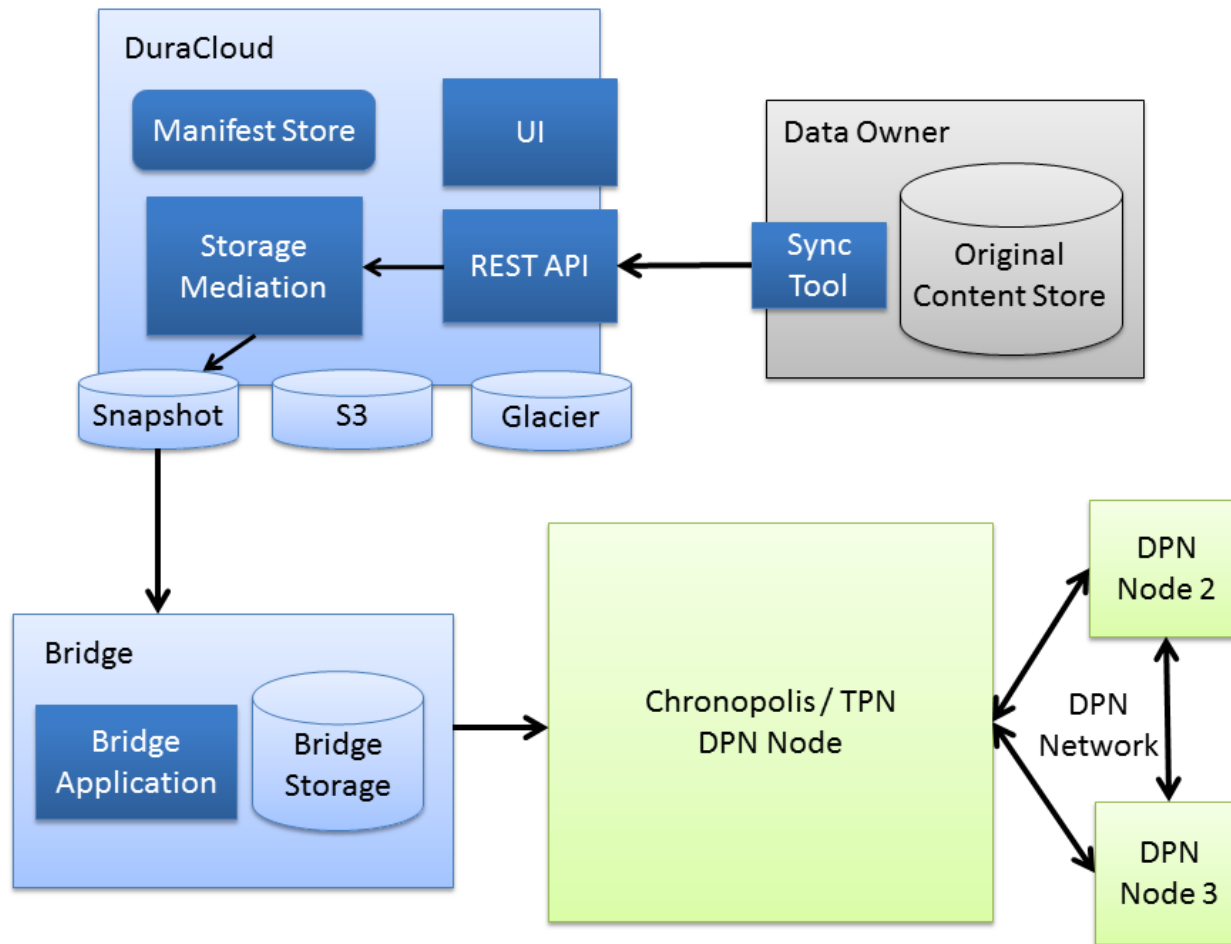
## DuraCloud Vault

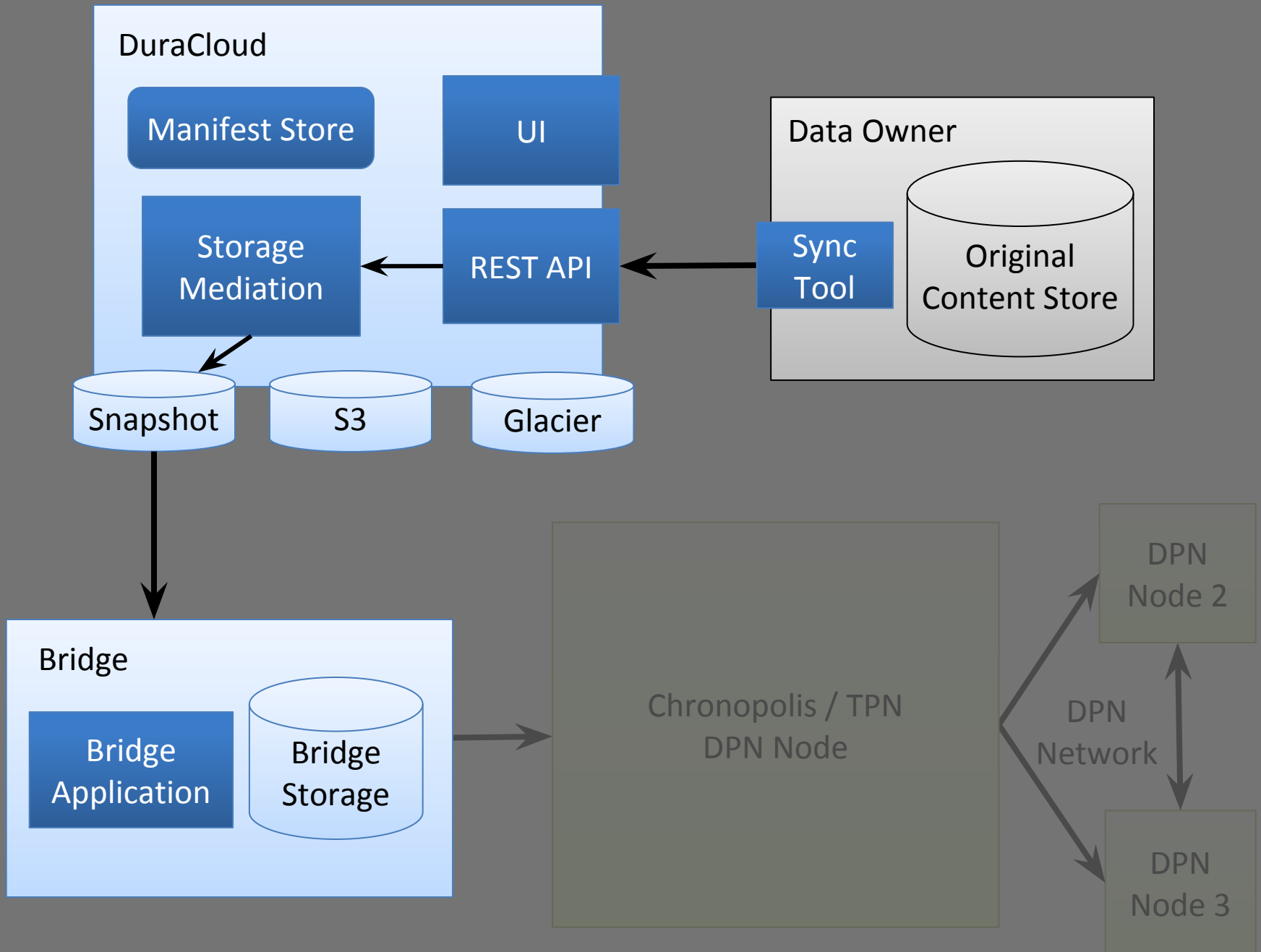
- Front end: DuraCloud
- Back end: Chronopolis

## Texas Preservation Node - TPN

- Front end: DuraCloud @ TDL
- Back end: TACC (Texas Advanced Computing Center)

# Integration Strategy





# Chronopolis

---

Sibyl Schaefer, Digital Preservation Analyst, UCSD

# Chronopolis

Digital preservation storage network spanning multiple institutions and geographic regions

Focused on: *active* preservation – constant checking of items

First ingest date: 2008

Trusted Digital

Repository Certification: 2012



# Partnering with DuraCloud

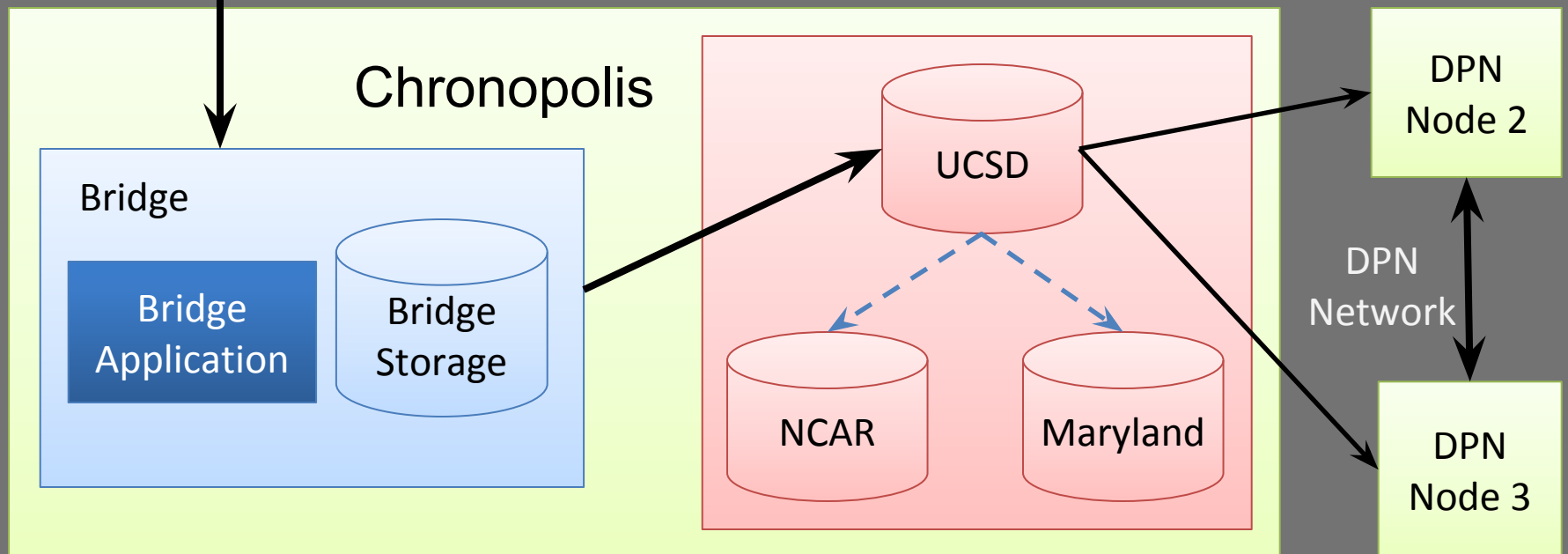
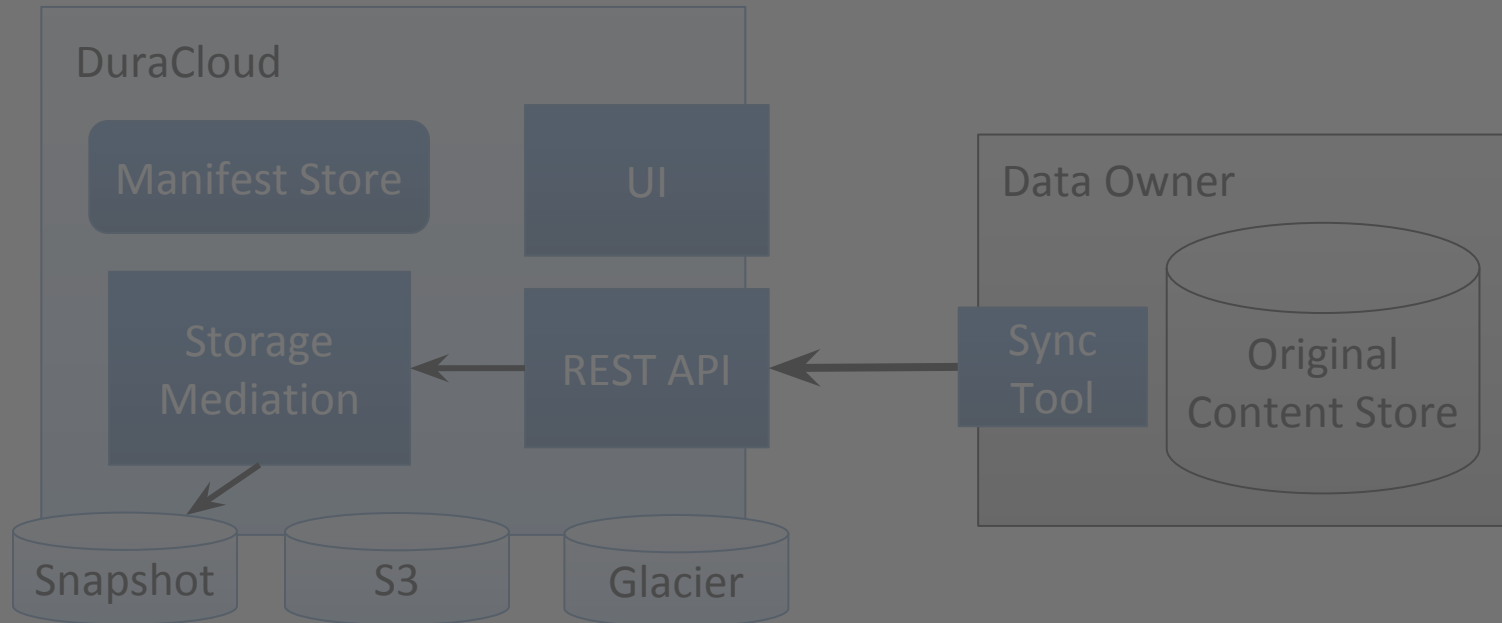
- **DuraCloud as a pathway to Chronopolis**
  - Provides an existing hosted user interface, reduced need for new development
  - Simplifies the process of moving content in and out of the systems
  - Shared institutional values
- **Chronopolis as a storage provider for DuraCloud**
  - Extends the DuraCloud network to include a non-commercial, highly distributed, dark archive option

# DuraCloud Vault



## **Providing an end-to-end DPN node solution.**

- Agnostic to content format, type, and size
  - All data is welcome
- Full access to content through DuraCloud
- Only DPN node offering ingest to any DPN member at DPN soft launch (July 1, 2015)
- <http://duracloud.org/duracloud-vault>



# Chronopolis backend

- Content pulled from staging, then verified
- Copies created in one of the distributed data centers (option to distribute to all three)
- Coordination with DPN to push content to nodes within DPN



Bridge

Ingest service

Ingest

Postgres

Replication service

JBOD

NCAR

ACE

Replication service

Isilon

UCSD

ACE

Replication service

JBOD

UMIACS

ACE



---

Ryan Steans, Director of Operations

# Texas Preservation Node Partners

A Cooperative Project



**Texas Advanced Computing Center** – High Performance super computer center with 6 PB available within just “Corral”



**UT Libraries** – Providing oversight of the project and supporting DPN development



**TDL** – Developing ingestion point for users of TPN via DuraCloud™



# TPN First Users

## TDL Membership



# Why DuraCloud?

- Open Source
- Flexible enough to handle multiple user requirements and extendible
- DuraCloud fit with TDL's use of existing DuraSpace technology (DSpace)
- Before DPN - DuraCloud experiments already underway between TDL and TACC
- Able to build upon work done with Chronopolis with different TPN architecture
- Extend local preservation plan to include DPN

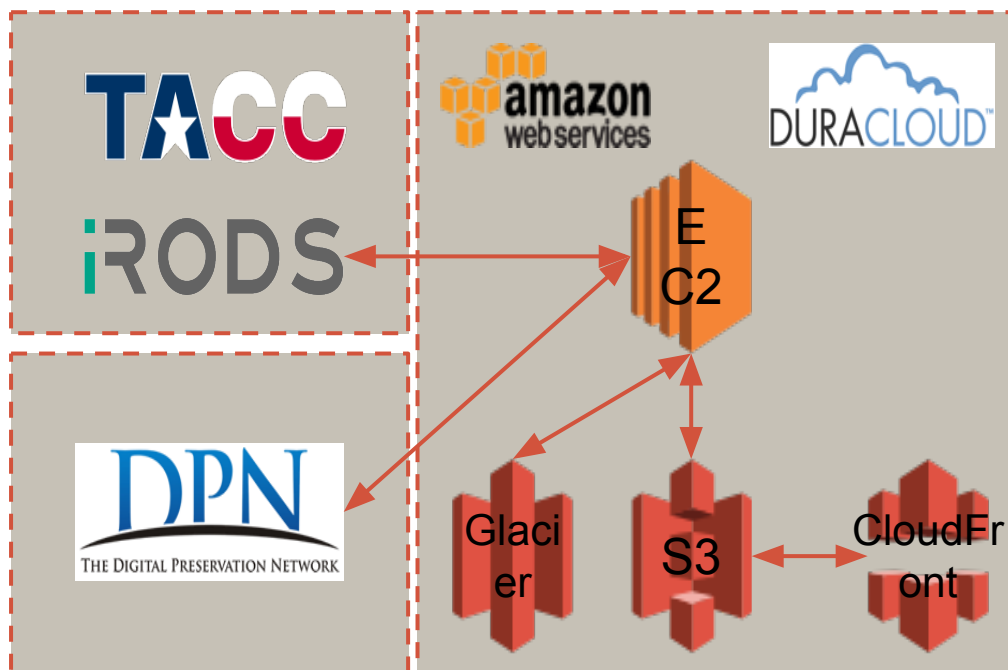
# DuraCloud™@TDL as a Service

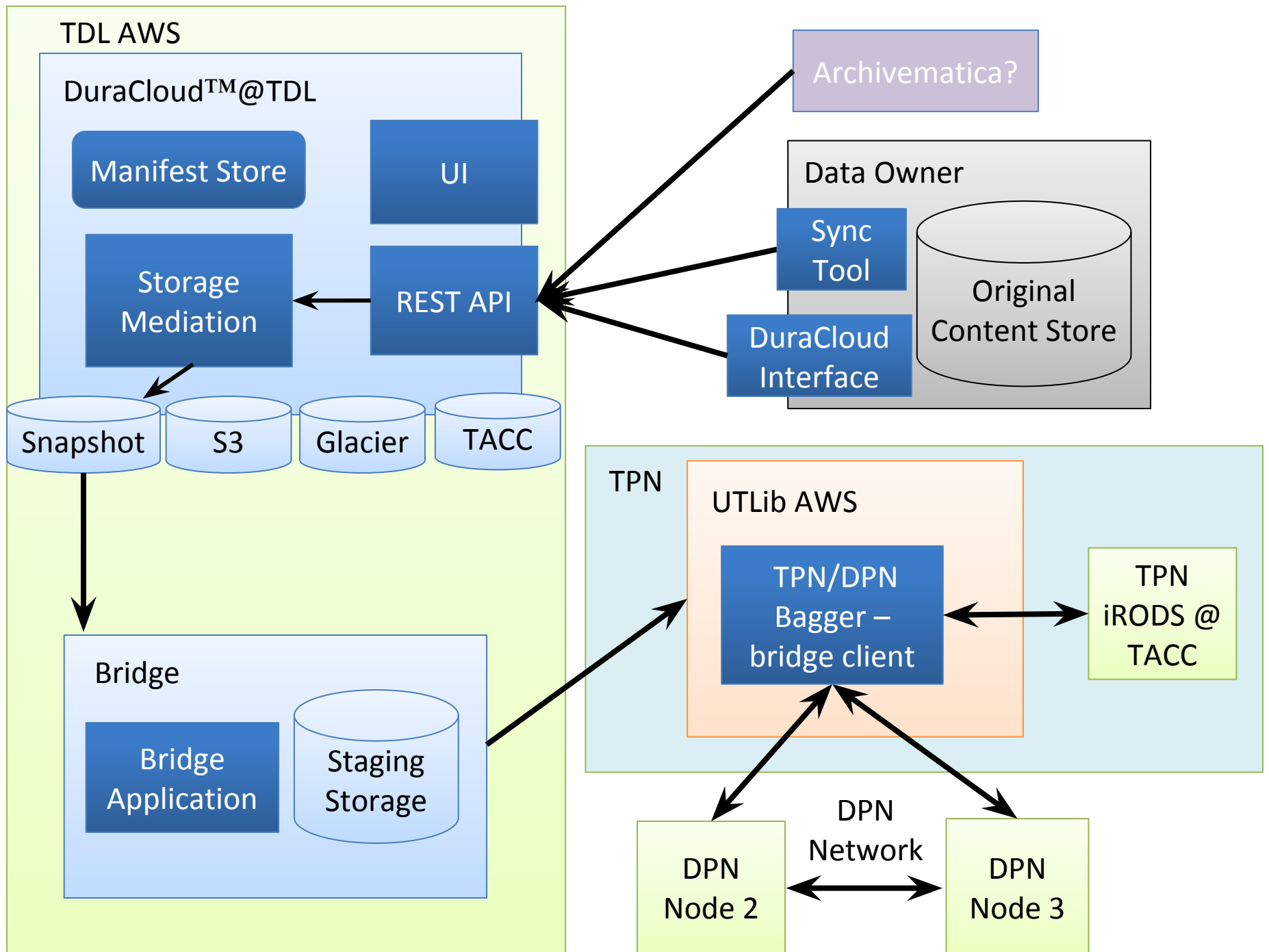
- Available as of January 2015 (it's here!)
- Consistent Upgrades - at DuraCloud 3.2
- Charge-back model – Members pay for what they use
  - Cost varies depending upon storage option
  - TDL Membership covers cost of running the basic service, technical support, etc...



# TDL Preservation Architecture

- EC2 Computing
- S3
  - Content Delivery Network
- Glacier
  - “Dark” Storage
- CloudFront
  - Streaming
- iRODS @ TACC
  - “Dark” Storage
- TPN/DPN





# Contributions

## DuraCloud

- <https://github.com/duracloud/duracloud>
- <https://github.com/duracloud/snapshot>
- <https://github.com/duracloud/mill>

## University of Maryland Institute for Advanced Computer Studies

- <https://github.com/msmorul/irods-api>
- <https://gitlab.umiacs.umd.edu/adapt/ace>

## Texas Digital Library

- <https://github.com/TexasDigitalLibrary/duracloud>
- <https://github.com/TexasDigitalLibrary/irods-api>

# GitHub



# Questions?

Bill Branan

[bbranan@duraspace.org](mailto:bbranan@duraspace.org)

Sibyl Schaefer

[sschaefer@ucsd.edu](mailto:sschaefer@ucsd.edu)

Ryan Steans

[rsteans@austin.utexas.edu](mailto:rsteans@austin.utexas.edu)