

Chronopolis and DuraCloud: Doing integration right

PASIG Spring 2015

David Minor
Chronopolis



Bill Branan
DuraSpace

DuraCloud

- Hosted digital preservation service
- Storage mediation system
- Preservation services
- Simple pathway to cloud-based digital content storage
- Open source - built to support the library and archives community

Chronopolis

- Digital preservation storage network spanning multiple institutions and geographic regions
- Focused on long-term digital content preservation
- System based on *active* preservation – constant checking of items

Why integrate? Common goals.

- Digital content preservation is important to the future of society
- All preserved digital content should be handled equally
- Need to provide an economically viable option to support the preservation needs of all institutions, regardless of size or technical capability
- Need to simplify the preservation process as much as possible

Each organization has focused on different parts of the service spectrum

Integration

- **DuraCloud as a pathway to Chronopolis**
 - Provides a hosted user interface and tooling which simplifies the process of moving content
- **Chronopolis as a storage provider for DuraCloud**
 - Extends the DuraCloud network to include a non-commercial, highly distributed, dark archive option

... in service to DPN

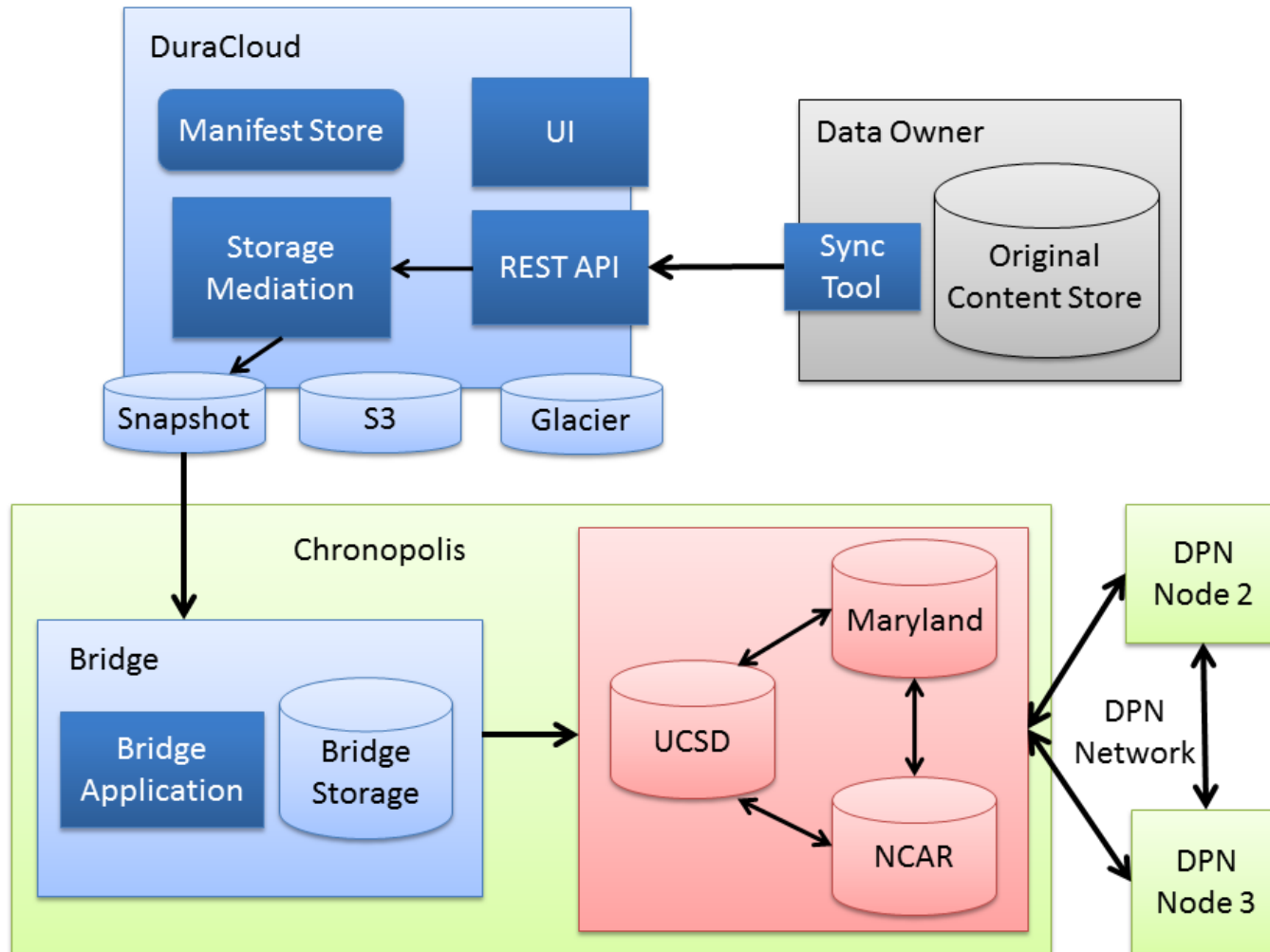
Providing an end-to-end DPN node solution

- Agnostic to content format, type, and size
 - All data is welcome
- Provides an open and straight-forward pathway to storing content in DPN
- Extended DPN - full access to content through DuraCloud

The integration problem

- Two very independent existing systems with different workflows and processes
- DuraCloud works with real-time data, Chronopolis works with well defined data collections

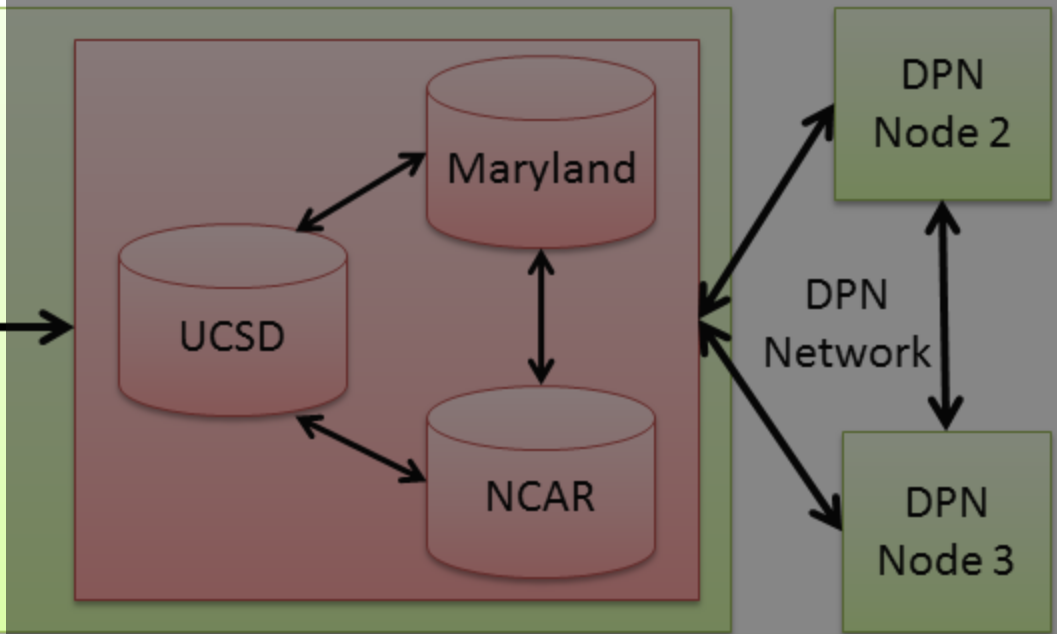
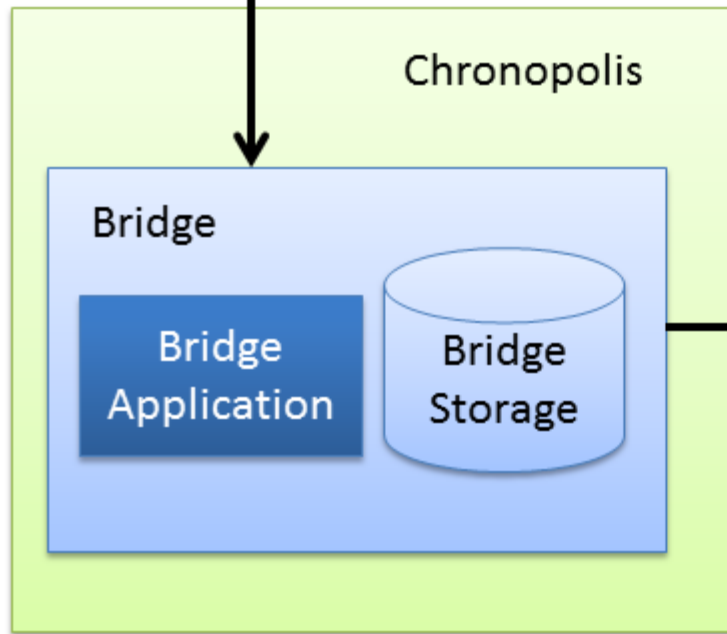
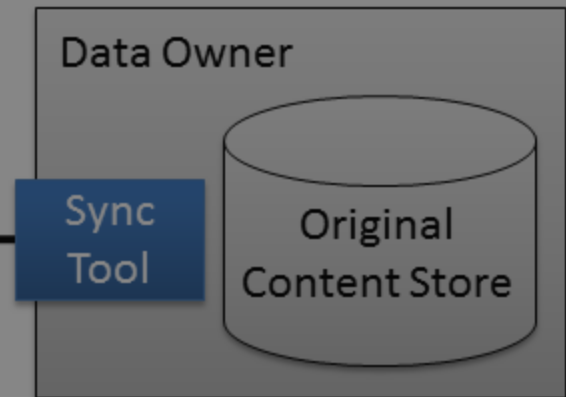
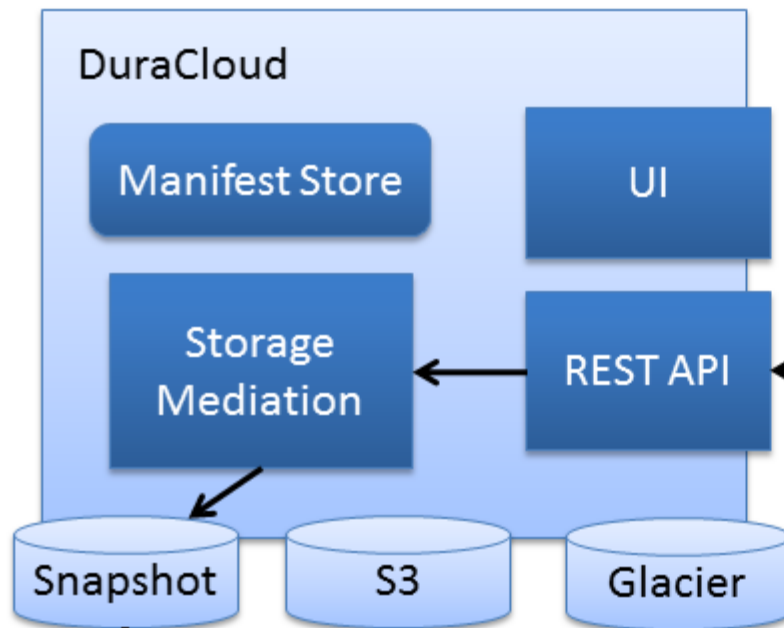
The integration solution



DuraCloud as the Front End

SyncTool and REST API provide a convenient means for transferring content

- Content remains online and accessible until the choice is made to perform a snapshot
- DuraCloud communicates with a "bridge" application

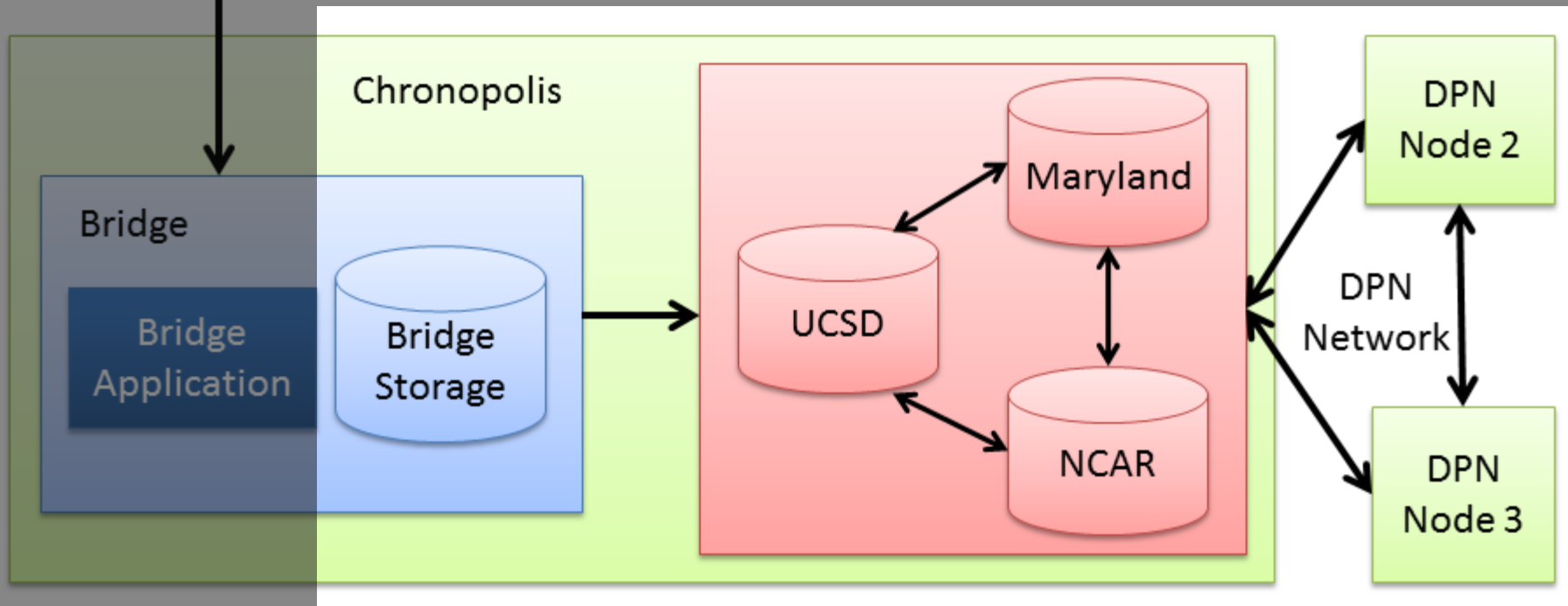
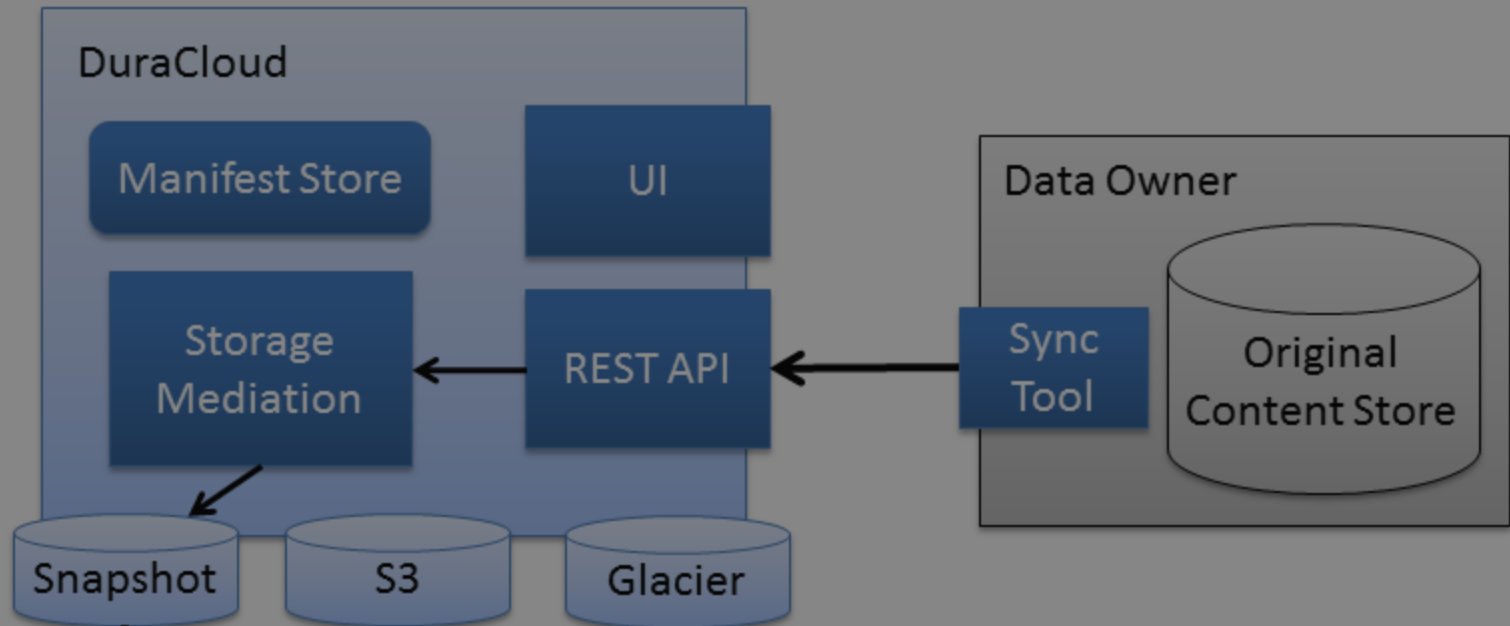


Snapshots

- Collections of content which are bundled together as a group for preservation
- Contents of the snapshot completely determined by user
- Snapshot content list and metadata remains visible in DuraCloud
- Snapshots can be restored back to DuraCloud, making content available for download

The Bridge

- Infrastructure software layer written to bridge the gap between the DuraCloud and Chronopolis systems
- Moves content from DuraCloud to Chronopolis staging - with verification via manifest along the way



Chronopolis backend

- Content pulled from staging, then verified and bundled
- Copies created in each of the distributed data centers
- Coordination with DPN to push content to nodes within DPN

Status

- Primary functionality is complete
- Initial DPN pilot with Dartmouth proved capability
- Secondary pilot scheduled to begin in April
- Production offering later this year

TakeAways

- Sometimes the best way to integrate two systems is to not require either system to know anything about the other
- DPN is one of the most exciting initiatives in the academic preservation space in many years
- Keep your ears open for the launch of the DuraCloud and Chronopolis service

Thank You

∞ Contact

- David Minor: dminor@ucsd.edu
- Bill Branan: bbranan@duraspace.org

∞ Web sites

- Chronopolis: chronopolis.ucsd.edu
- DuraSpace: www.duraspace.org
- DuraCloud: www.duracloud.org
- DPN: www.dpn.org