



The DuraCloud Workshop

Your hosts: Bill Branan & Carissa Smith

What we plan to accomplish.

1. The Basics
2. Lab: Getting Started with DuraCloud
- 3. Break @ 10:30 (in Cosmopolitan AB)**
4. Architecture and Integrations
5. Lab: DuraCloud Deep Dive
- 6. Lunch @ 12:30 (in Cosmopolitan AB)**

The (un)format.

- Informal, let's have some fun :)
- Ask questions!
- What do you want to learn about DuraCloud?

A little background.

- First, there was DuraSpace:
 - 501(c)(3) not-for-profit
 - created from Fedora Commons, Inc. + DSpace Foundation
 - “committed to our shared digital future”
- DuraSpace = Projects + Services
 - DSpace, Fedora, VIVO
 - DuraCloud, DSpaceDirect, ArchivesDirect

Then there was DuraCloud.

1. Library of Congress funding 7/14/2009
 - a. 2009 market assessment & initial development
 - b. 2009 pilot phase 1 (storage)
 - c. 2010 pilot phase 2 (interface, services)
2. DuraCloud Launched 11/1/2011
3. Integrations/Features 2012-2013
4. Chronopolis/DPN 2014-2015

This is DuraCloud.

- open source software
- offsite storage
- one interface
 - multiple cloud copies
 - multiple storage providers
- automated synchronization
- audited content health checks
- consolidated billing and agreements
- integrations with other services
 - DSpace, Archive-It, Archivematica, etc.

This is **not** DuraCloud.

- a repository replacement
- an Amazon account billed by DuraSpace
- intelligent about metadata
- a network drive or file system

The burning question.

- Why DuraCloud over Amazon?
- Reality: DuraCloud is built on Amazon (but with additional preservation services):
 - MD5 content health checks
 - content upload tools
 - synchronization (local to cloud & cloud copies)
 - integration with other storage providers
- Honest Truth: sometimes it makes sense to use DuraCloud & other times Amazon directly

So...

- ❑ Use DuraCloud when you want...
 - ❑ multiple copies of your content
 - ❑ cloud storage with different providers
 - ❑ cloud storage in different geographic locations
 - ❑ simple tooling for managing data transfer
 - ❑ audited health checks of cloud provider
 - ❑ a pathway to DPN
 - ❑ integration with other open source technologies
 - ❑ one annual predictable storage invoice
 - ❑ personalized support

Let's get started.

1. Go to duracloud.org.
2. Click “no waiting: try it now”.
3. Create a trial account.
4. Check your email.
5. Log in at <http://trial.duracloud.org>.
6. Poke around.

Spaces

- filter
- examples
 - trial381
 - welcome

Content Items

Refresh

Showing 1 - 5 of 5

type prefix

DuraCloud-to-DSpace-Integration.pdf
Fox.jpg
Wildlife.mp4
examples/images/Mountains-Summer.jpg
examples/images/Mountains-Winter.jpg

Space Detail

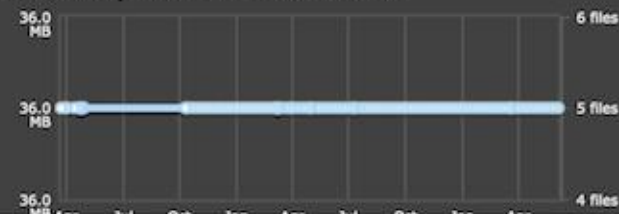
examples

Details

Items	5	Recount
Created	2013-03-20	
Streaming Host	s3pgizgw4hi7o.cloudfront.net	

History

Cumulative Byte and File Counts Over Time



Let's kick the tires.

1. Upload a file or two.
2. Copy a file.
3. Delete a file.
4. Add properties and tags.
5. Navigate between storage providers.

Spaces

[+ Add Space](#)

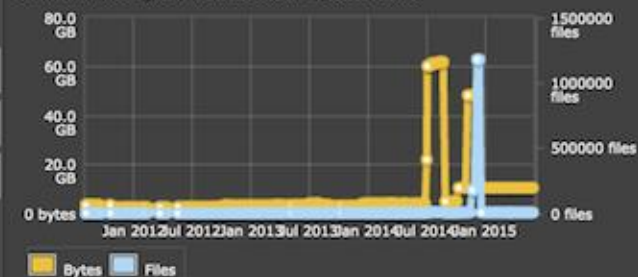
filter

- ☐ bb-space
- ☐ bb-test
- ☐ carissa-folder-test
- ☐ carissa-images
- ☐ carissa-upload-test
- ☐ carissa-uploads
- ☐ carissa-video-test
- ☐ carissa-video-test-viewer
- ☐ db-test
- ☐ downloads
- ☐ dspace-dev-backup
- ☐ dspace-dev-trash
- ☐ dspace-test-backup
- ☐ dspace-test-trash

Spaces

History

Cumulative Byte and File Counts Over Time



Spaces as of May 29, 2015

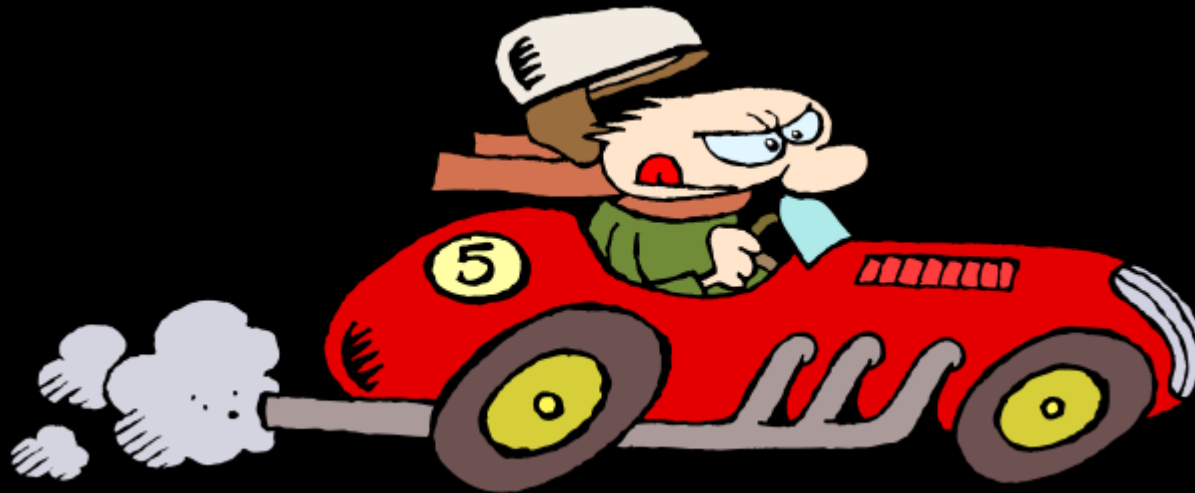
[Data](#)

I have the power.

- Create, edit, delete spaces.
- Make content “public”.
- Turn streaming on/off.
- Manage user/group permissions.
- View full account storage details.

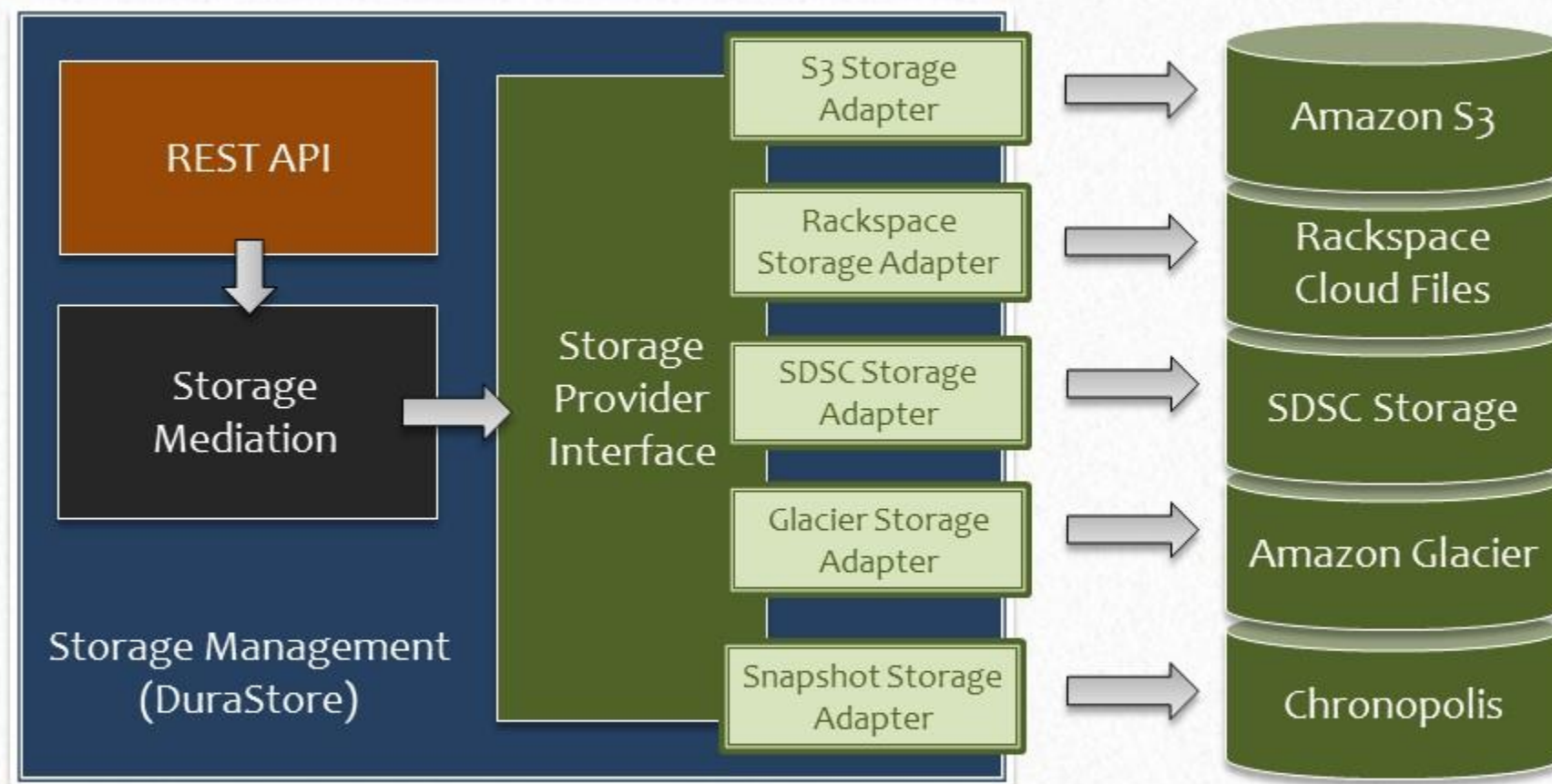
Break time!

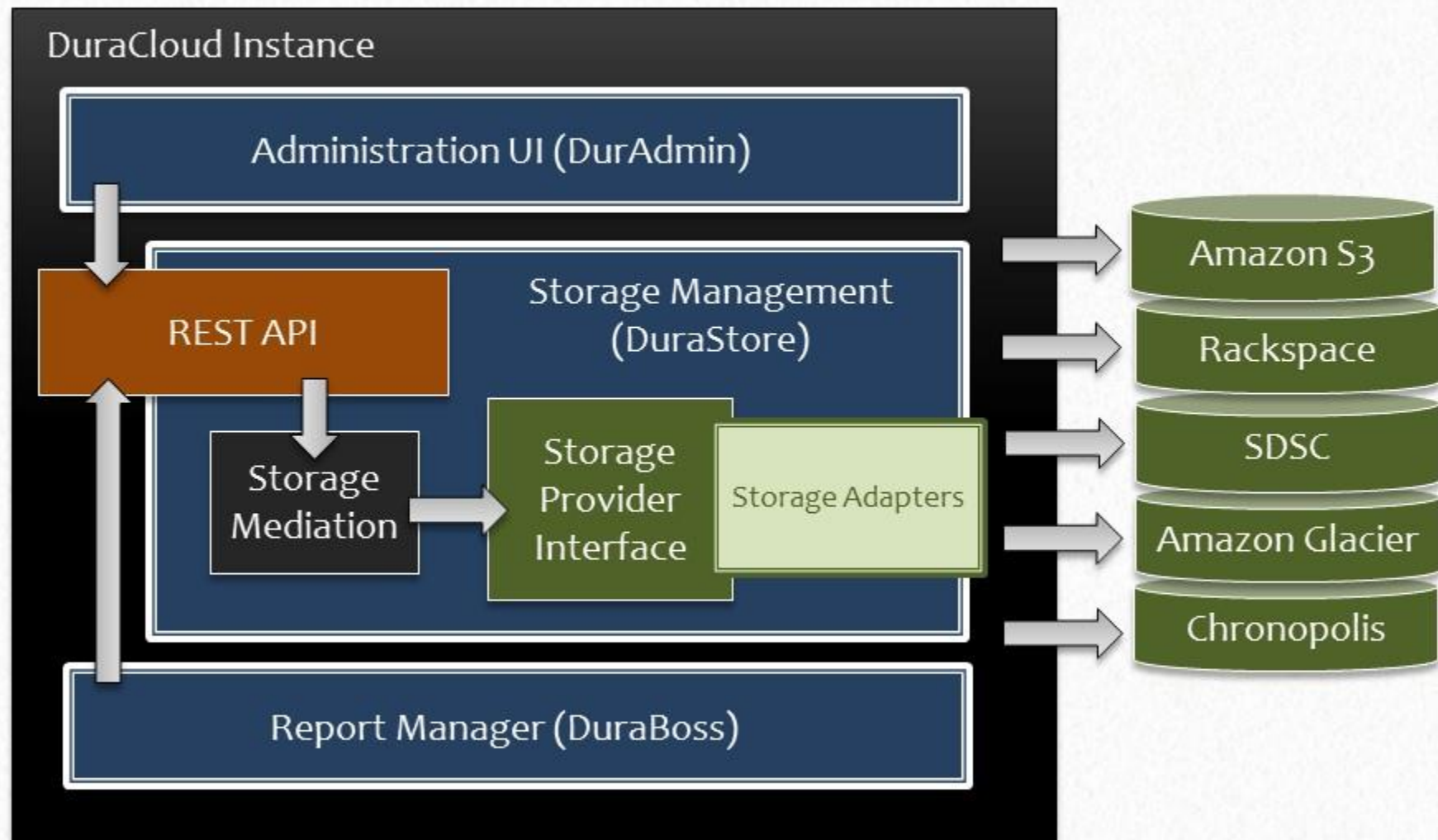
See you back here at 11:00am.

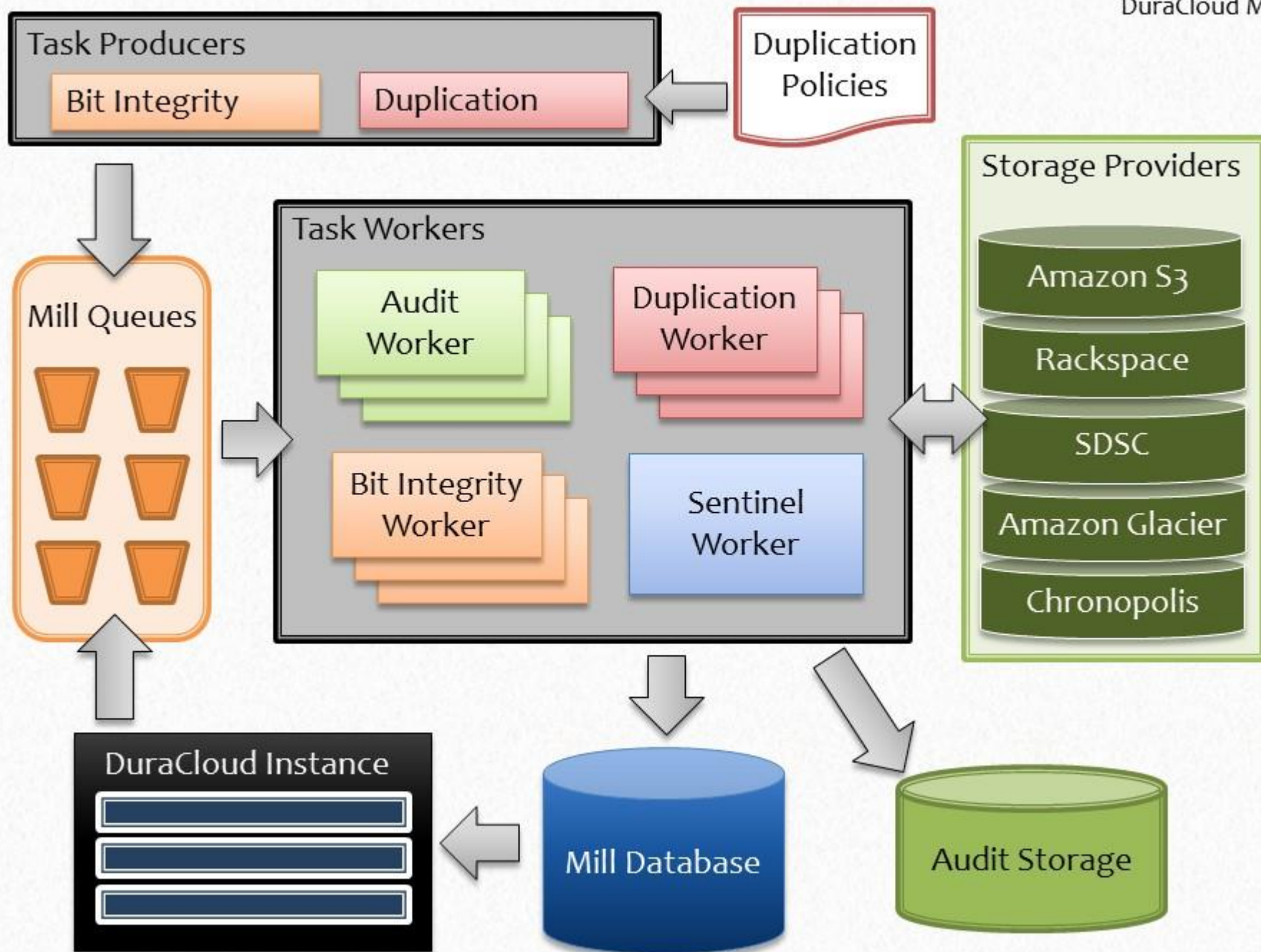


The architecture.

- Components
 - DuraStore - storage systems, the interface to all storage providers
 - DurAdmin - web-based user interface
 - DuraBoss - manages storage reporting
 - The Mill - handles auditing, manifest generation, bit integrity, and duplication
 - Standalone tools - synchronization, retrieval
- REST API







Standalone Tools

- Sync Tool
 - Tool for transferring local content to DuraCloud
 - Option of UI-based or command-line interface
 - Built-in transfer-rate optimization utility
 - Lots of options for selecting files and directories and specifying how you want the files stored
- Retrieval Tool
 - Command line tool for retrieving content that is stored in DuraCloud space
 - Can retrieve a content listing, all content in a space, or a specific set of desired content

Let's upload more content.

1. Log in to <http://trial.duracloud.org>.
2. Click on your trial space.
3. Click the “Upload” button.
4. Click “[Want to make uploads automatic?](#)”
5. Download the synchronization utility (or as we call it “the sync tool”).

Welcome!

Welcome to the DuraCloud Sync Tool

Once set up, this application will automatically back up your vital digital files to DuraCloud, thus ensuring your data remains accessible and secure.



Continue

So many integrations.

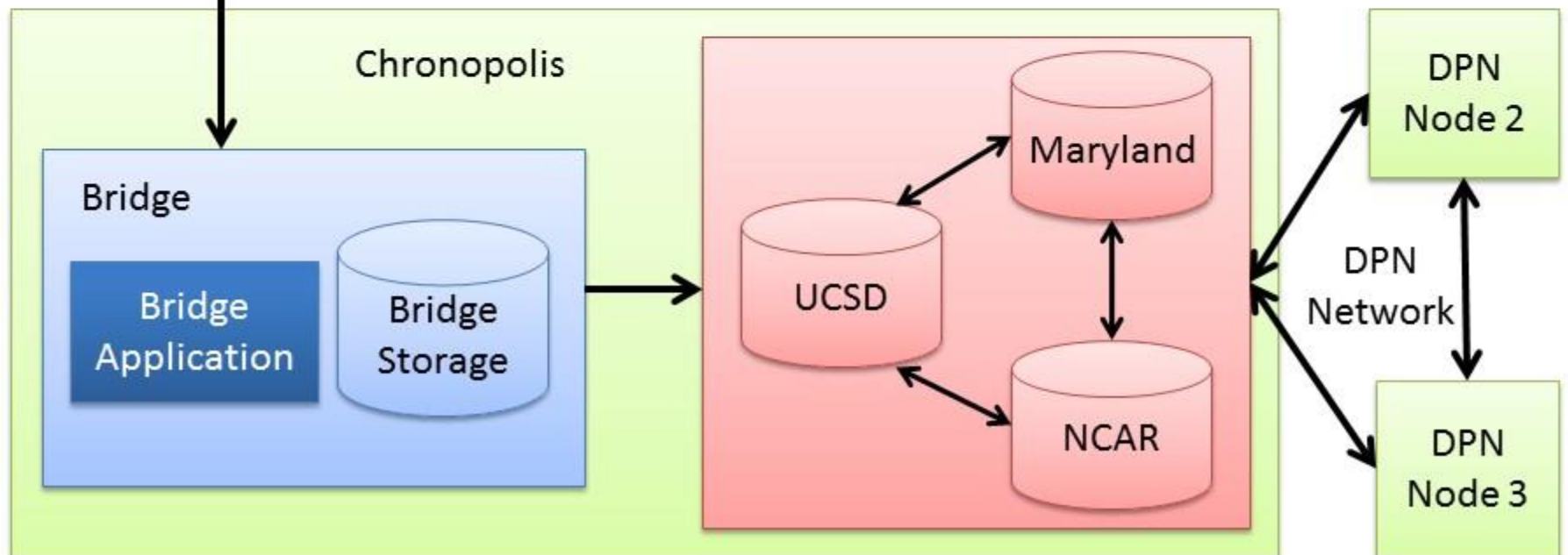
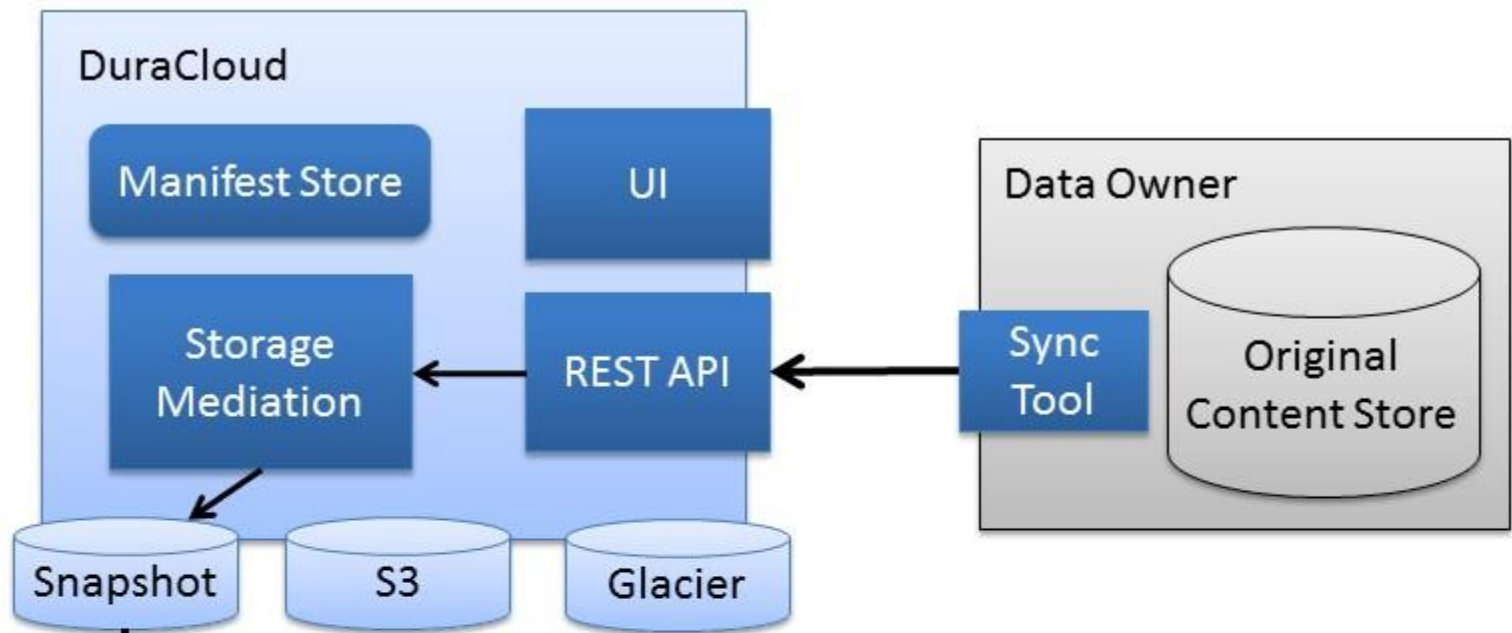
- Chronopolis/DPN
- DSpace/DSpaceDirect
- Archivematica/ArchivesDirect
- Archive-It
- Upcoming: Fedora, Figshare

DuraCloud + Chronopolis

- Chronopolis = digital preservation storage network spanning multiple institutions and geographic regions
 - Comprised of UC San Diego, NCAR, and Univ of Maryland
 - Dark archive
 - Focused on long-term digital content preservation
- Extends the DuraCloud network to include a non-commercial, highly distributed, dark archive option

DPN

- DuraCloud provides a pathway to Chronopolis
- Chronopolis provides a pathway to DPN
- DPN = Digital Preservation Network, a nascent national preservation system
- DuraCloud Vault:
 - DuraCloud + Chronopolis = DPN node
 - <http://duracloud.org/duracloud-vault>
- TPN (Texas Preservation Node) - another DPN node, also using DuraCloud



DuraCloud + DSpace


- DSpace = institutional repository
- DSpace generates AIPs
- DSpace Curation Task System configured to transfer new content to DuraCloud
- Can also perform complete batch backup of all stored data to DSpace
- Also used to manage restores

<https://wiki.duraspace.org/display/DSPACE/ReplicationTaskSuite>

DSPACE DIRECT

- Hosted and managed DSpace repository
- Simple to start-up
- Wide selection of available features
- Free version upgrades
- Migration support
- Automatic backup to DuraCloud included at no additional cost
- www.dspacedirect.org

DSpace Curation Task Setup

 **DSpaceDirect Demo**

Profile: Demo Administrator | [Logout](#)

Home → Curation Tasks

System Curation Tasks

Handle of DSpace Object:

Hint: Enter [your-handle-prefix]/0 to run a task across entire site (not all tasks may support this capability)

Choose from the following groups:

[DuraCloud Backup Tasks](#) ▾

Task:

- ✓ Estimate Storage Space for AIP(s)
- Read Odometer
- Transmit AIP(s) to Storage
- Verify AIP(s) exist in Storage
- Audit against AIP(s)
- Remove AIP(s) from Storage
- Restore Missing Object(s) from AIP(s)
- Replace Existing Object(s) with AIP(s)
- Restore Missing Object(s) but Keep Existing Objects
- Restore Single Object from AIP
- Replace Single Object with AIP

Search

[Go](#)

Browse

[All Content](#)

[Communities & Collections](#)

[By Issue Date](#)

[Authors](#)

[Titles](#)

[Subjects](#)

My Account

[My Exports](#)

[Logout](#)

[Profile](#)

[Submissions](#)

Administrative

[Control Panel](#)

[Access Control](#)

[People](#)

[Groups](#)

[Authorizations](#)

[Content Administration](#)

[Items](#)

[Withdrawn Items](#)

[Private Items](#)

[Import Metadata](#)

[Batch Import \(ZIP\)](#)

[Registries](#)

[Metadata](#)

[Format](#)

[Statistics](#)

[Curation Tasks](#)

[DSpace software](#) copyright © 2002-2013 [DuraSpace](#)

[Contact Us](#) | [Send Feedback](#)

Powered by **DSpaceDirect**

A service from **DURASPACE**

DuraCloud + Archivematica

- Archivematica = workflow tool for generating preservation ready AIP packages
 - Customizable OAIS-based workflow processing
 - Permanent ID generation
 - Checksumming
 - Virus checking
 - File format identification and validation
 - Metadata extraction and generation
 - Normalization
- DuraCloud as preservation storage

archivesDIRECT

- Hosted Archivematica with pre-configured DuraCloud backup
- Launched in 2015!
- www.archivesdirect.org



DuraCloud + Archive-It

- Archive-It = web archiving service from the Internet Archive
- Content automatically copied from Archive-It to DuraCloud as a secondary backup
- Provides Archive-It users with direct download access to their entire collection of WARC (web archive) files

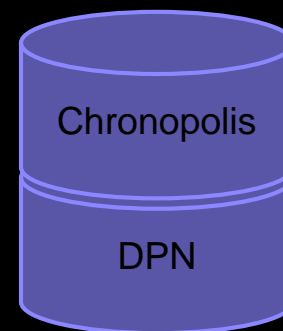
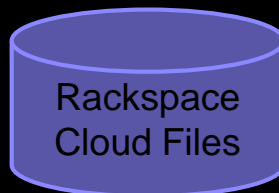
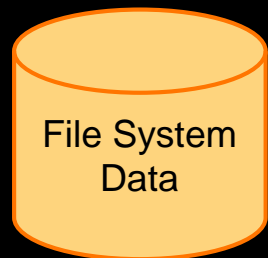
On the horizon



- DuraCloud Vault release
- Figshare integration
- Fedora 4 integration

archivesDIRECT

SPACE DIRECT



The REST API

Time to get our hands dirty!



<https://ac...requester/>

Lab tips

- All urls will begin with:
<https://trial.duracloud.org> (https)
- Use the correct request type (GET, PUT, POST, HEAD, DELETE)

<https://wiki.duraspace.org/display/DURACLOUDDOC/DuraCloud+REST+API>

Thank you!

- Contact
 - Bill Branan bbranan@duraspace.org
 - Carissa Smith csmith@duraspace.org
- User Group
 - duracloud-users@googlegroups.com
- YouTube
 - www.youtube.com/duracloudvideos