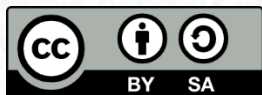




Hydra Partners Meeting March 2012

Bill Branan
DuraCloud Technical Lead



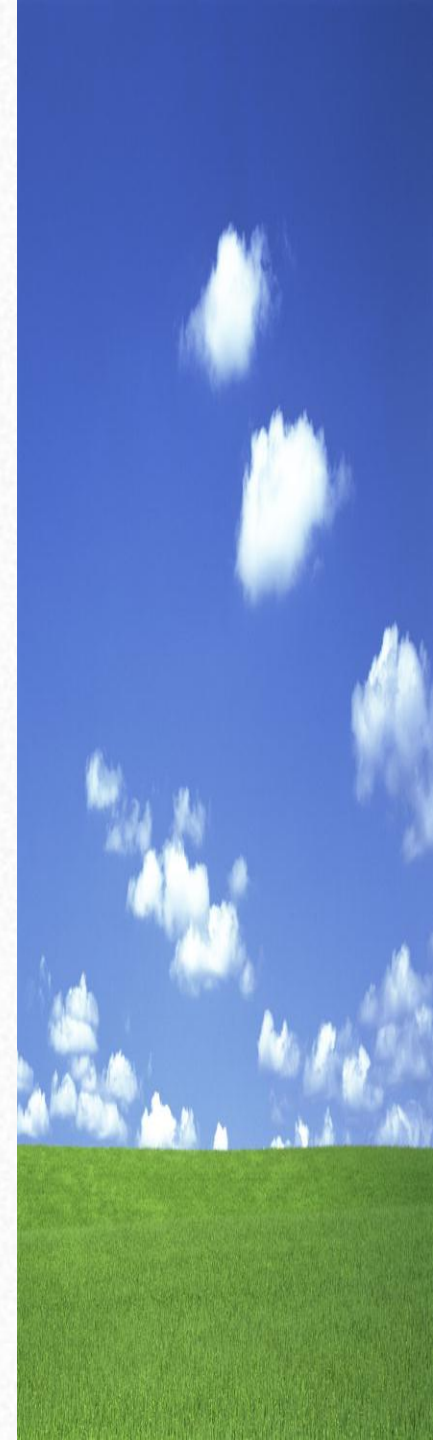
Agenda

- Introduction
- Architecture
- Running DuraCloud
- Cloud Gotchas
- Initiatives
- DuraCloud for Research
- Integrations



Introduction

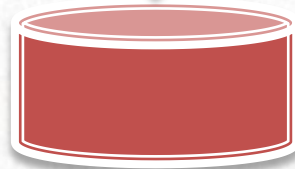
- DuraCloud is:
 - Hosted service
 - Runs on cloud-based compute systems
 - Connects to cloud-based storage systems
 - Provides a service platform
 - Open source software suite
- Primary Goals:
 - Simplify the path to the cloud
 - Add preservation sensibilities to the cloud
 - Provide a service platform
 - Enable a community-based cloud solution



Unique URL:
<http://yourinstitution.duracloud.org>



Organization/Administrator using
DuraCloud



Repository, Content
Management System, File
System
(via web, REST APIs, utilities)



*DuraCloud platform running
on a compute cloud*

Programming interfaces

UI
System
Administration

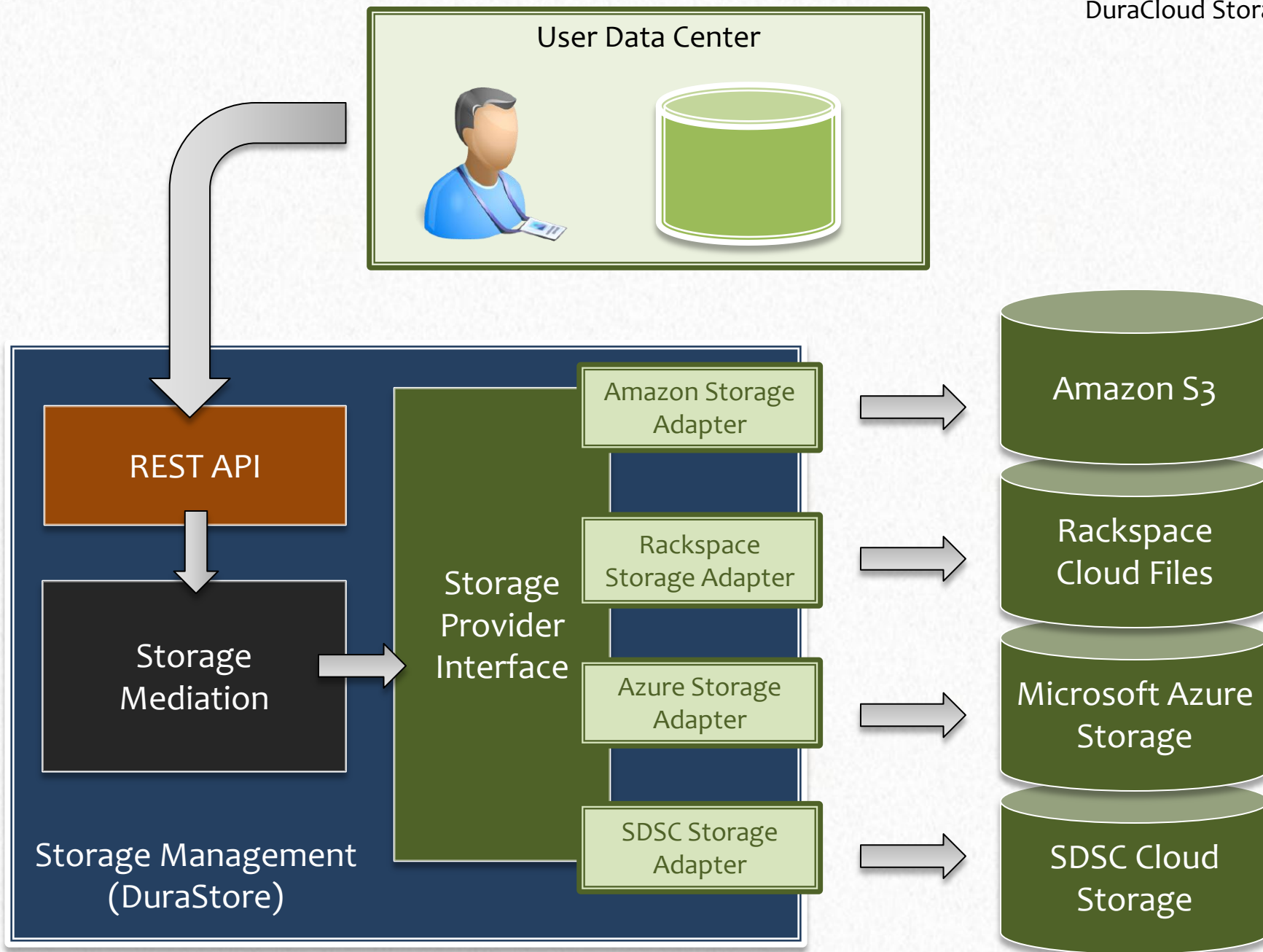
API
Reporting &
Automation

API
Storage
Management

API
Service
Management



Services:
Backup
Health check
Synchronization
Replication



Storage REST Interface

Space Actions

- Add Space
- Get/Set Space Properties
- Get Spaces List
- Get Space Content List
- Get/Set Space Access
- Delete Space

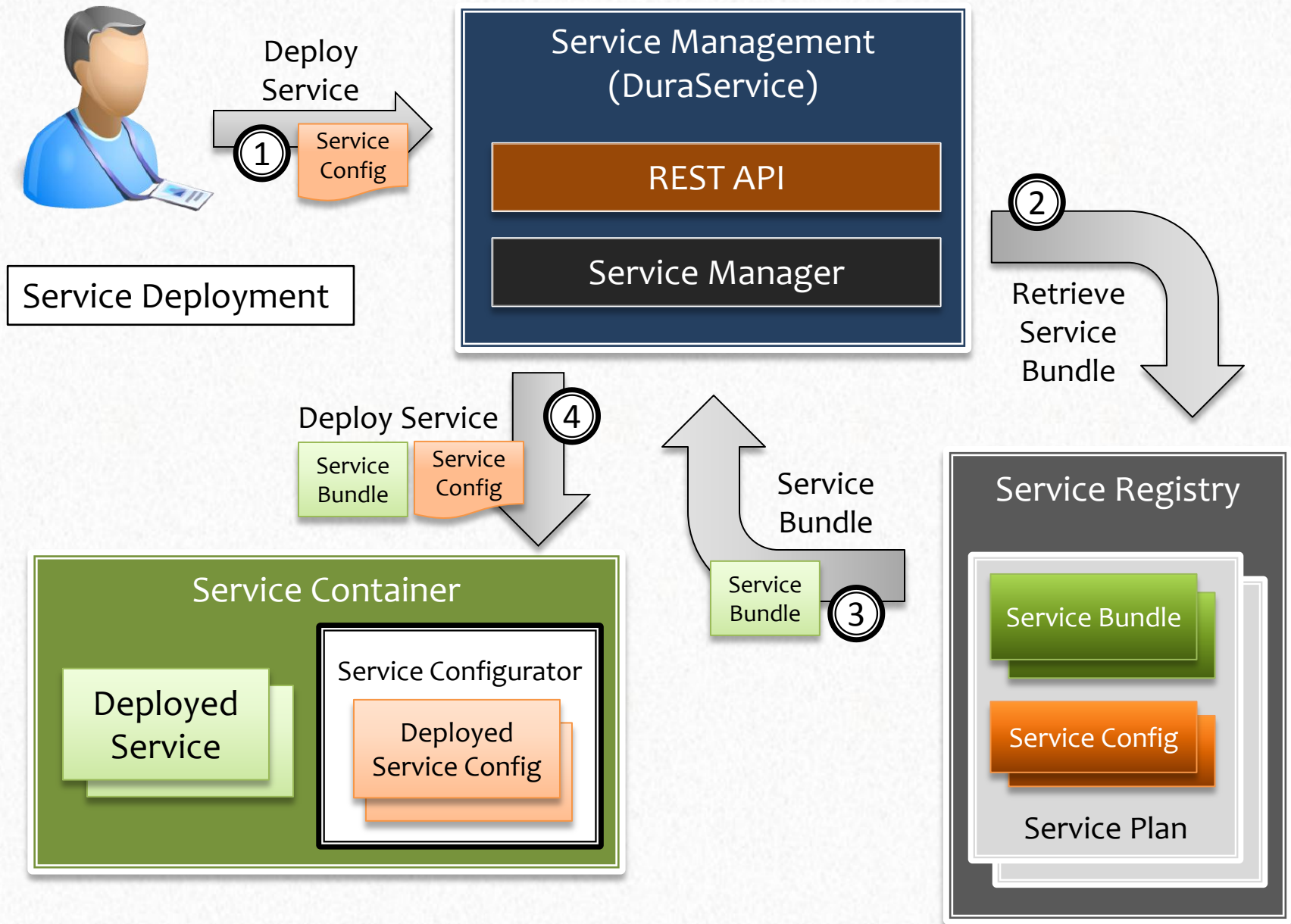
Content Actions

- Add Content
- Get/Set Content Properties
- Get Content
- Copy Content
- Delete Content

Storage Provider Interface

Other Actions

- Get Stores
- Get Tasks List
- Perform Task



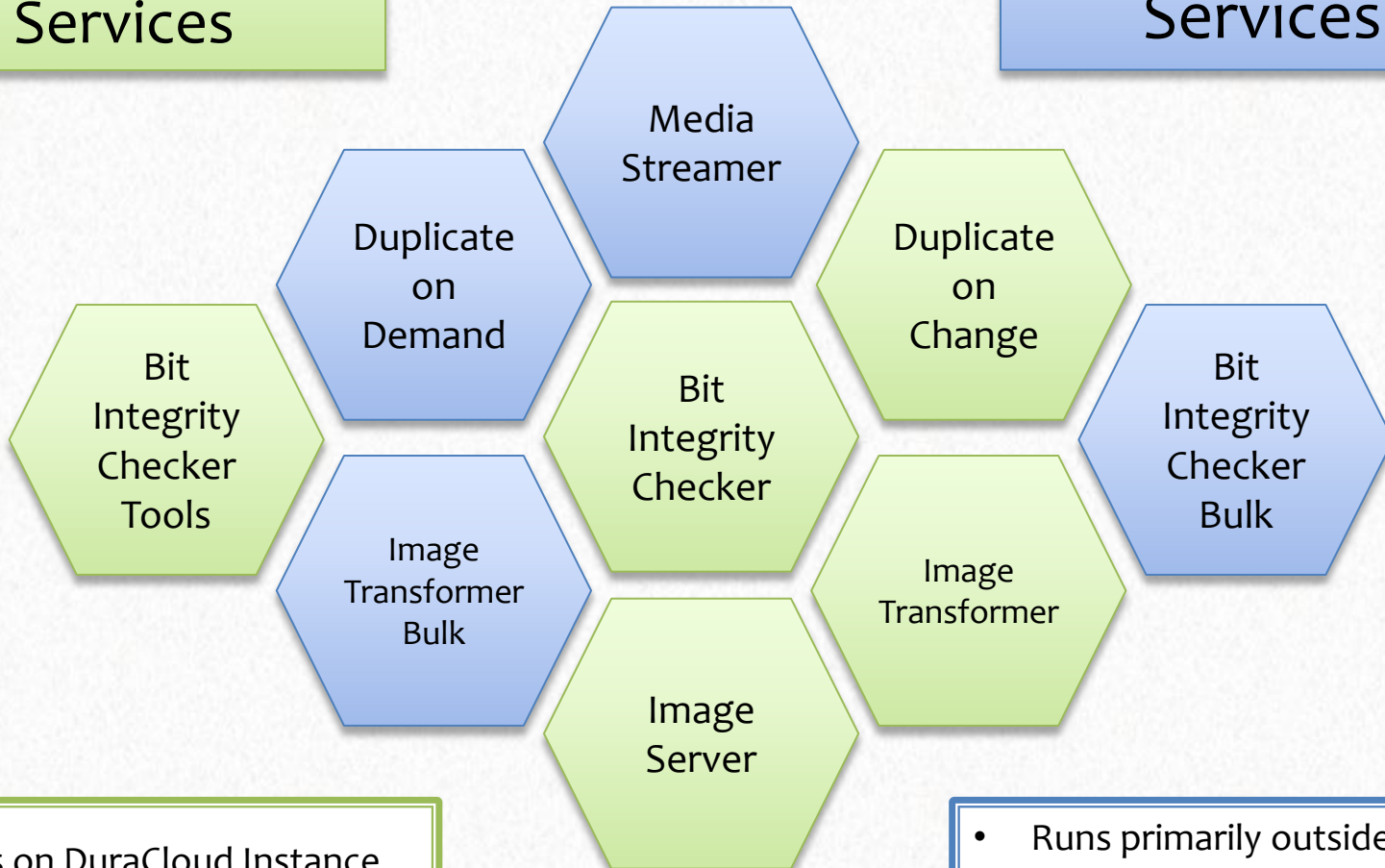
Service REST Interface

- Get Services
- Deploy Service
- Get (Deployed) Service
- Get Deployed Service Properties
- Update Service Configuration
- UnDeploy Service



DuraCloud Instance Services

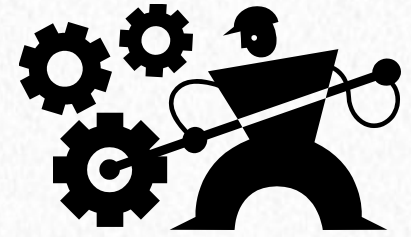
DuraCloud Distributed Services



- Runs on DuraCloud Instance
- Connects to DuraStore
- Can be direct Java service
- Can be deployed web app

- Runs primarily outside of DuraCloud Instance
- Connects to DuraStore
- Makes use of cloud network or computation features

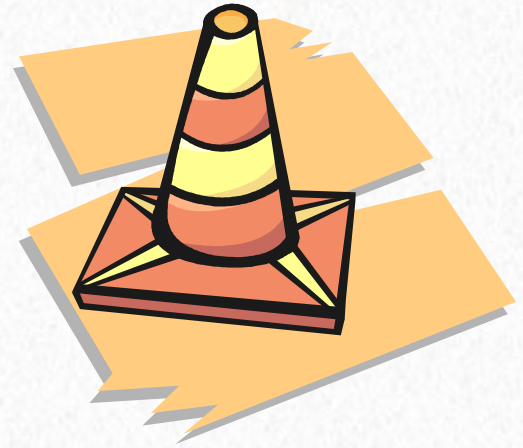
Running DuraCloud



- Set of four Java web applications
 - Deploy into a servlet container (Tomcat)
- OSGi container
 - Used to manage DuraCloud services
 - Pre-deployed dependency bundles
- Initialization
 - Connect to loud storage
 - Point to apps and services

Cloud Gotchas

- Eventual consistency
- Server volatility
 - Application “state”
 - Monitoring
- HTTP limitations
- Bandwidth limitations
- Bit integrity
- Storage provider APIs
 - Standards?



Current Initiatives

- Simplification
 - Service automation
 - Service and storage display integrated
- Multi-tenancy
- Internet2 Net+ Service
- Shibboleth integration
- DuraCloud for Research (DfR)

DuraCloud for Research (DfR)

- Grant funded by Sloan foundation
- Goals
 - Encourage the preservation of research data
 - Facilitate cooperation between researcher and institutional data managers
 - Provide tools and services to support the research process

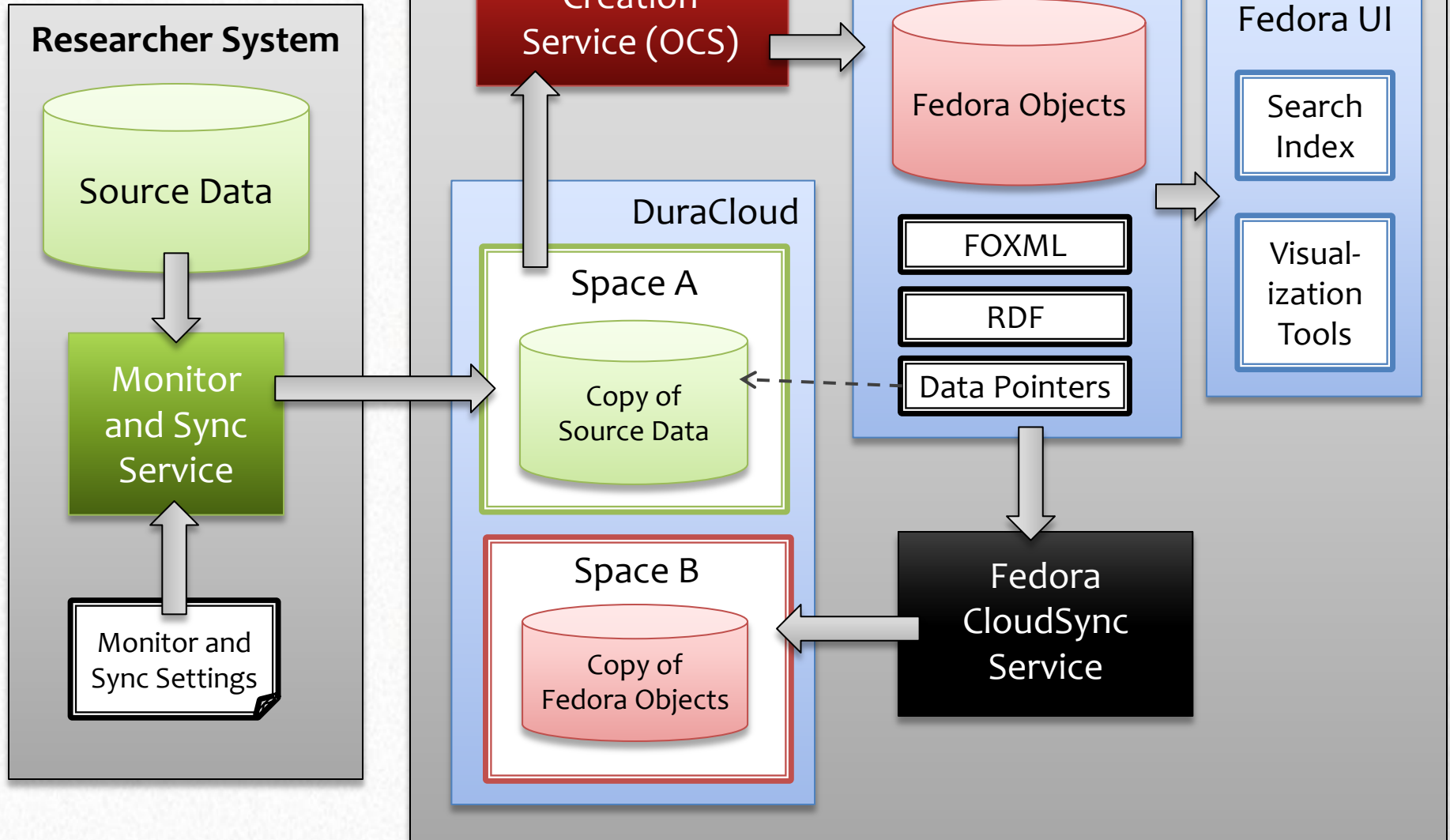
DfR Priorities

1. Connect the operational and archival phases of the data management lifecycle.
2. Create simple workflows across the data management lifecycle that automatically capture metadata and provenance. (... and create incentives for additional metadata creation)
3. Ensure confidentiality, security, privacy, and predictability of data in the cloud. (Trust and Control)
4. Automate basic metadata creation and “catalogue” creation.
5. Create interoperability of operational systems, archiving solutions, and discovery systems used by specific research communities.

DfR Principles

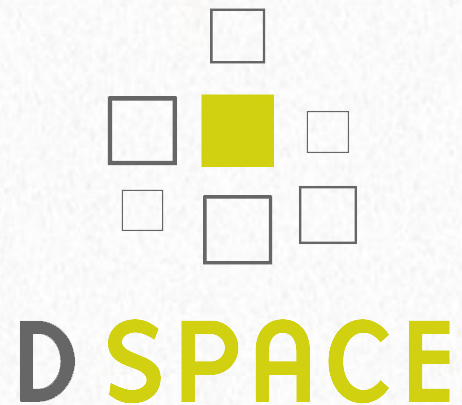
- Open source, enterprise software solution
- Capture data close to the source
- Don't interfere with researchers' processes
- Provide incentives, added value for metadata creation
- Easy to use; workflows for collaboration, hand-off to institution

DuraCloud for Research Architecture Sketch




DSpace + DuraCloud

- Add-On: Replication task suite
- Curation system tasks
- AIP packages
 - Collection, Community, or Repository
 - Multiple formats
- Estimate size, Transmit, Verify, Restore, ...



Dspace + DuraCloud

 mirage

[DSpace 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:

Replication Suite Tasks ▼

Transmit AIP(s) to Storage ▼

Dspace + DuraCloud



DURACLOUD™ ADMINISTRATOR

Dashboard Spaces Services

powered by **amazon** web services™

Spaces

☐

- ☐ my-dspace-backup
- ☐ x-duracloud-admin

Content Items

☐ Showing 1 - 10 of 10

<input type="checkbox"/>	COLLECTION@123456789-11.zip	
<input type="checkbox"/>	COMMUNITY@123456789-6696.zip	
<input type="checkbox"/>	ITEM@123456789-103.zip	
<input type="checkbox"/>	ITEM@123456789-6675.zip	
<input type="checkbox"/>	ITEM@123456789-6677.zip	
<input type="checkbox"/>	ITEM@123456789-6680.zip	
<input type="checkbox"/>	ITEM@123456789-6682.zip	
<input type="checkbox"/>	ITEM@123456789-6687.zip	
<input type="checkbox"/>	ITEM@123456789-6695.zip	
<input type="checkbox"/>	SITE@123456789-0.zip	<input type="button" value="Download"/> <input type="button" value="Delete"/>

Dspace + DuraCloud

Profile: Tim Donohue | Logout



[DSpace 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:

Search DSpace

[Advanced Search](#)

Browse

- [All of DSpace](#)
 - ♦ [Communities & Collections](#)
 - ♦ [By Issue Date](#)
 - ♦ [Authors](#)
 - ♦ [Titles](#)
 - ♦ [Subjects](#)

[My Account](#)

Dspace + DuraCloud



Profile: Tim Donohue | Logout



[DSpace Home](#) ♦ [Collections](#) ♦ [Curator](#)

Curate Collection: Test Files Collectionafsadfsadf

[Edit Metadata](#) [Assign Roles](#) [Content Source](#) [Curate](#)

Choose from the
following groups:

Replication Suite Tasks ▼

Task:

Audit against AIP(s) ▼

Perform

Queue

Return

Search DSpace

[Advanced Search](#)

Browse

- [All of DSpace](#)
 - ♦ [Communities & Collections](#)
 - ♦ [By Issue Date](#)
 - ♦ [Authors](#)
 - ♦ [Titles](#)
 - ♦ [Subjects](#)

Fedora + DuraCloud

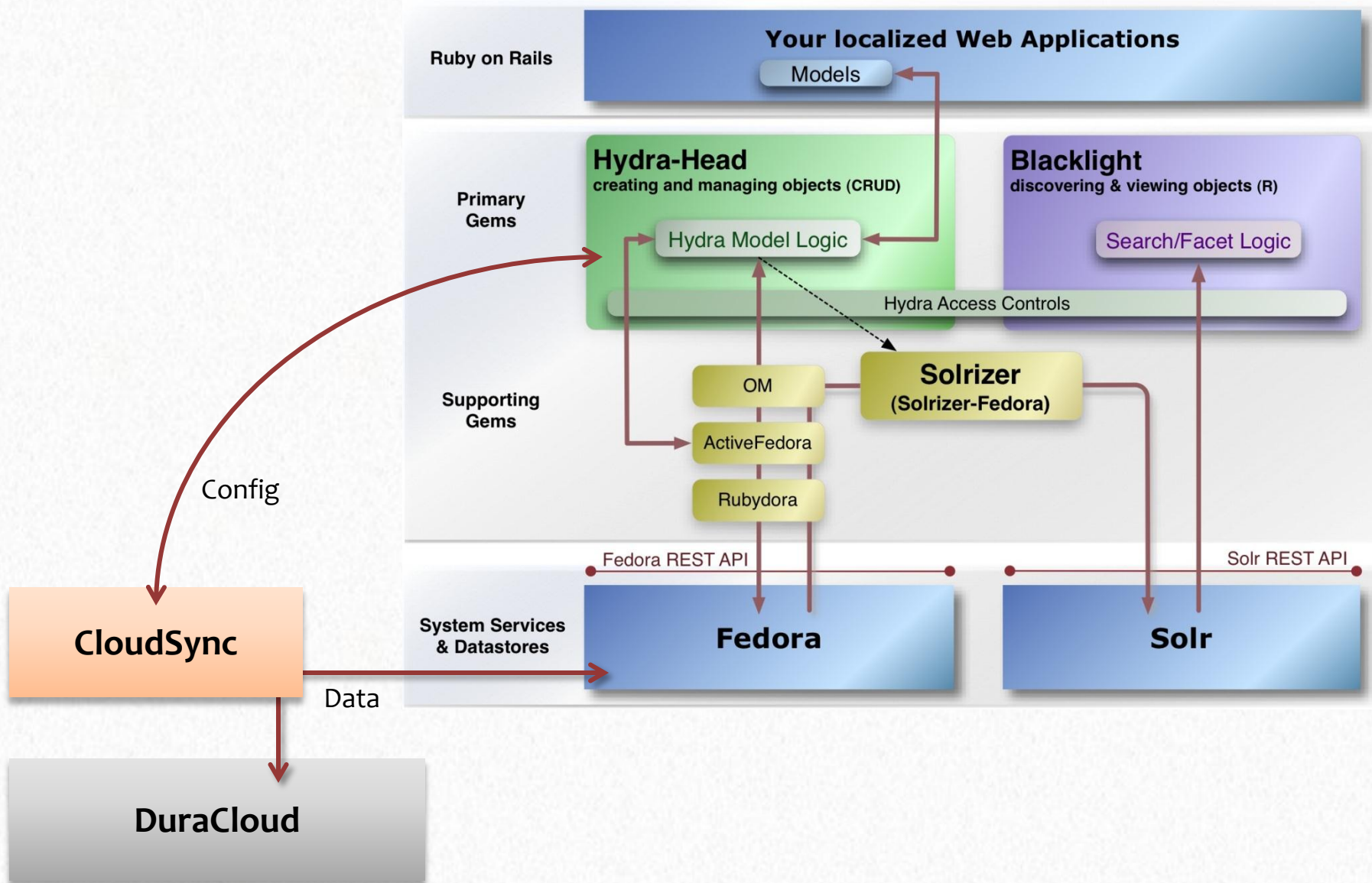
- Direct Akubra ✖
- Fedora CloudSync
 - Point to Fedora
 - Point to DuraCloud
 - Configure datasets
 - Perform Backup
 - Perform Restore



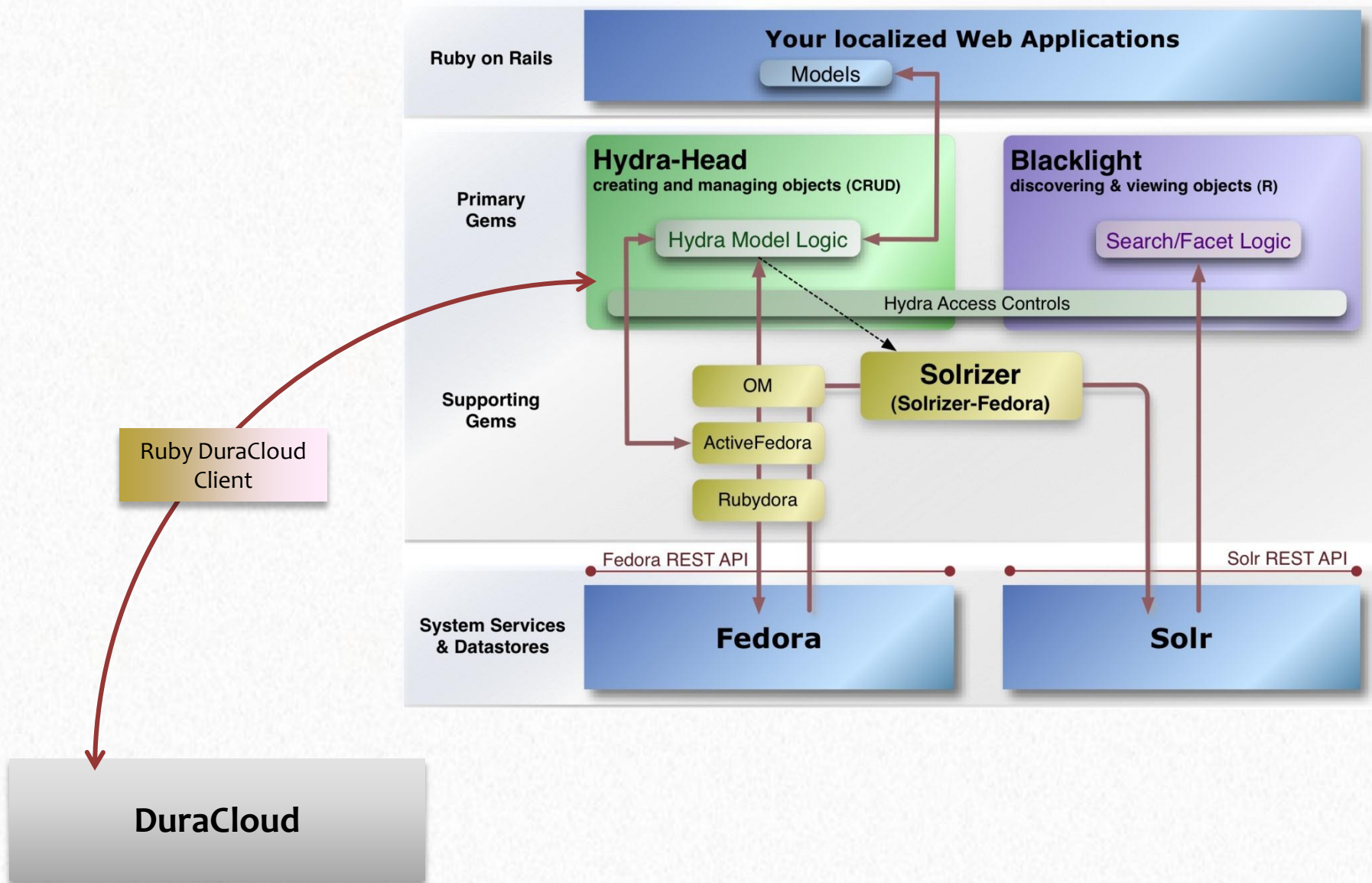
Comparison

	Typical Solution	CloudSync + DuraCloud
Backup Granularity	All Objects	Any Subset
Restore Granularity	All Objects	Any Subset
Skill Level	Advanced	Basic
Downtime	Yes	No
Offsite	DIY	Yes

Hydra + DuraCloud



Hydra + DuraCloud



Thank You!

bbranan@duraspace.org