

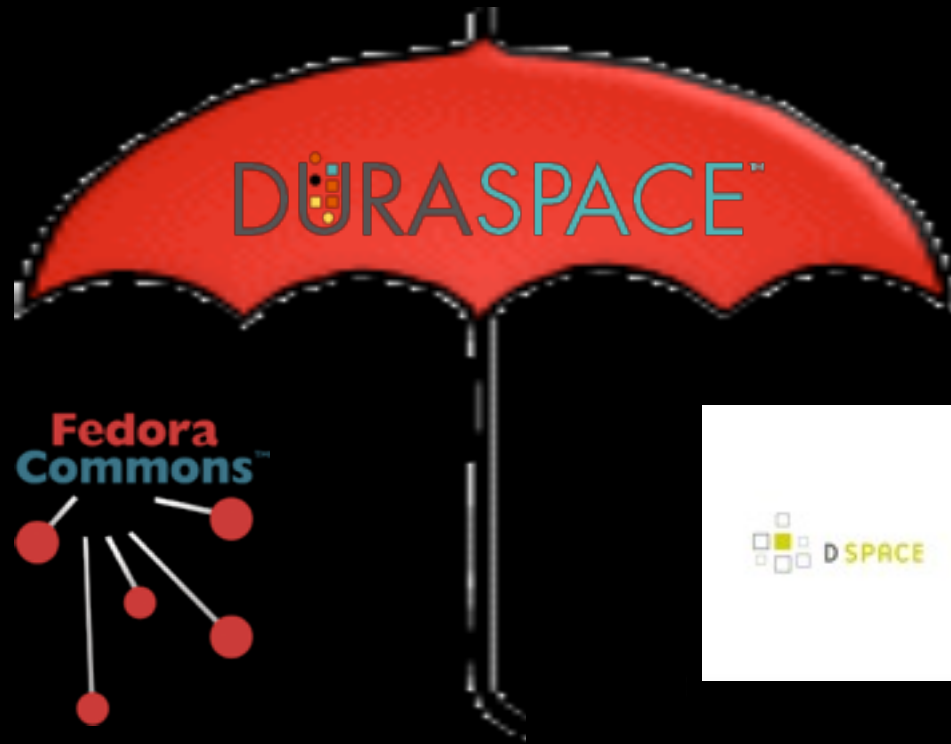
# DuraCloud

*Federated Repositories and Cyberinfrastructure  
Open technologies and services for  
managing durable data in the cloud*

The DuraSpace logo, identical to the one in the top left corner, featuring the words "DURA" and "SPACE" in grey and teal respectively, with a trademark symbol and a vertical bar of colored squares to the left.

Bradley McLean, CTO DuraSpace

# Open Source Portfolio



**DuraCloud**



# Goals of DuraSpace

## Stewardship:

Support and align open source development communities for DSpace and Fedora

## Innovation:

Think beyond existing platforms

New strategies for enabling access and preservation of digital content

## Sustainability:

Develop business model to sustain the non-profit and open technologies we support

## DSpace and Fedora Installations

Universities  
Research Centers  
Libraries  
Archives  
Cultural Heritage  
Government  
More...



Largest share of open repositories worldwide  
... over 700 institutions tracked in our registries

# Challenges (From our communities)

*Digital preservation and archiving is hard to achieve , even just **basic replication***

***Easy** and **elastic** provisioning of shared infrastructure (also across institutions!)*

*Robust compute environments for **data mining** and **analysis** of large datasets*

*Making digital **content** more **accessible** and **useable** to researchers*

# Implications for our future work



more distributed

more collaborative

more web-oriented

more open

more interoperable

# What About the Cloud?

A style of computing where massively scalable IT-related capabilities are provided “as a service” using Internet technologies to multiple external customers.  
(Gartner, 6/08).

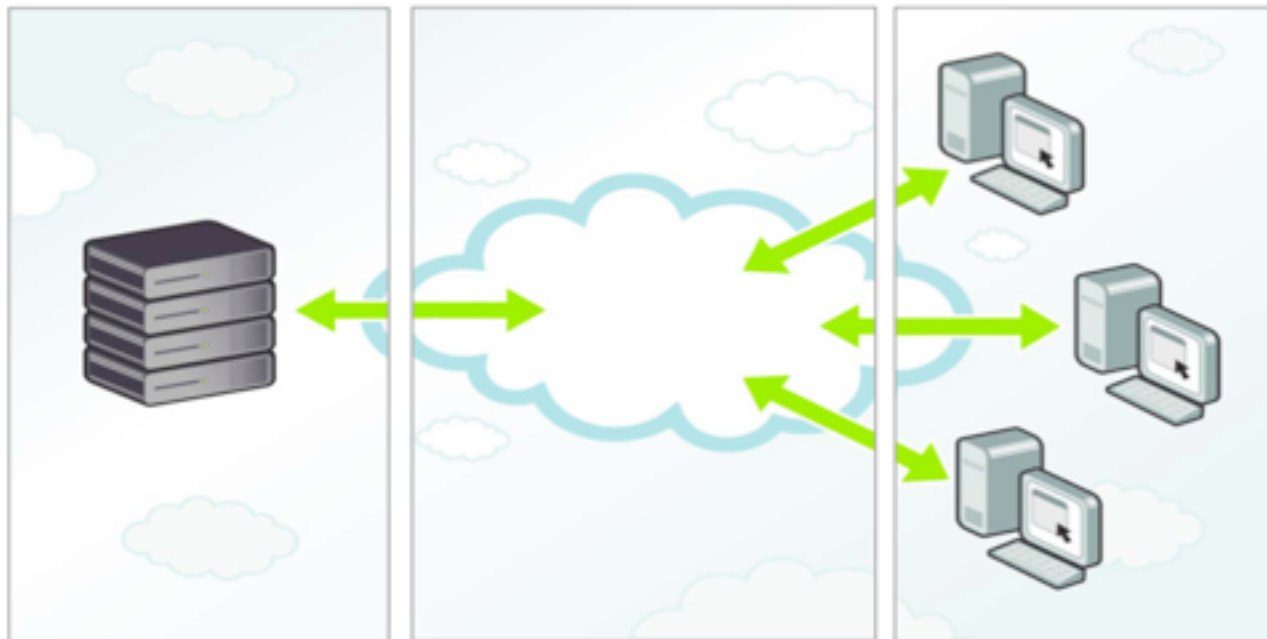


# Cloud services

• Software as a Service

• Infrastructure as a Service

• Platform as a Service





# Public Cloud Services

*Elastic web-based infrastructure for storage and compute*



# Economies of Scale and Cost

Public cloud providers drive cost down through scale, location and virtualization technology

Technology*	Cost Medium Datacenter	Cost Large Datacenter
Network	\$95 per Mbit/sec/ mo	\$13 per Mbit/sec/ mo
Storage	\$2.20 per Gbyte/mo	\$.40 per Gbyte/mo
Admin	140 servers/admin	>1000 servers/ admin

Large Datacenters (tens of thousands of computers)  
 Medium Datacenters (thousands)

Source: Hamilton, Internet-Scale Service Efficiency,, LADIS Workshop (Sept 08)

# Study of 605 government IT

**MeriTalk**  
The Government IT Network

**Cloud Benefits**

With Federal budgets shrinking, the cloud may offer an opportunity to save without sacrifice

*What are the most significant benefits of cloud computing for your agency?\**

<b>57%</b>	Reduced hardware requirements
<b>45%</b>	Reduced costs – pay-as-you go model
<b>35%</b>	Reduced staff requirements – allows for focus on more critical tasks
<b>33%</b>	Flexibility
<b>24%</b>	Potential for group collaboration
<b>22%</b>	Continuity of Operations Planning (COOP)

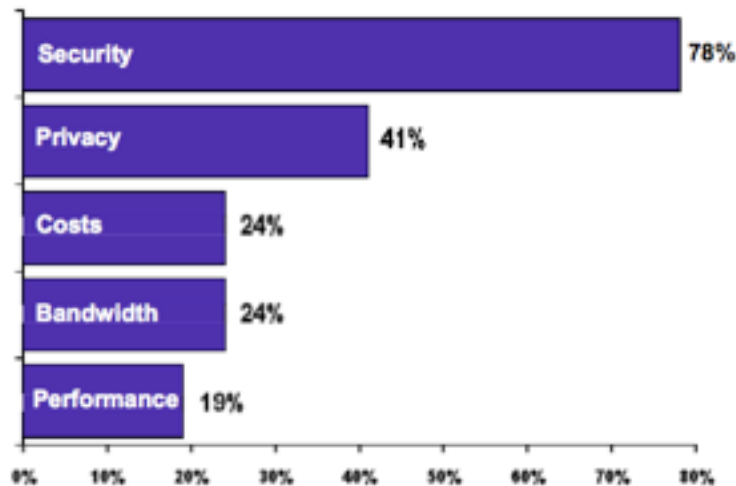
While Federal and industry IT managers agree on the top three benefits of cloud computing, Federal managers are twice as likely to view COOP as a top benefit – 22% to 11%

***Yet, only 13% utilizing cloud compute today***

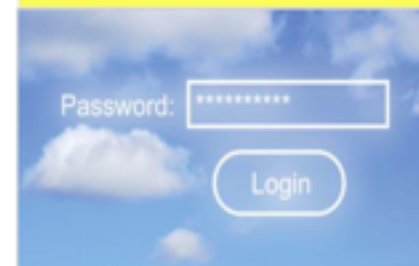
# Barriers

However, Federal IT managers see security concerns as a major hurdle to cloud advancement

*What are the biggest barriers to cloud computing within your agency?\**



Though Federal and industry IT managers both see it as the top barrier, Federal agencies are more likely than industry organizations to experience implementation challenges with security – 73% to 48%



# Here to stay

Despite uncertainty, Federal IT managers see the cloud as more than the latest tech trend

### Confident Outlook:

- 76%** Of Federal IT managers believe the cloud is "here to stay"
- 61%** Believe that in five years, the majority of large enterprises will rely on cloud computing for core applications

Federal IT managers are more likely than industry managers to hold this optimistic outlook – 61% to 51%

### Aggressive Timeline:

- 2.6** Average number of years that Federal IT managers believe it will take for their agency to realize the benefits of cloud computing



<http://www.meritalk.com/2009-cloud-consensus.php>

# DuraCloud Proposition

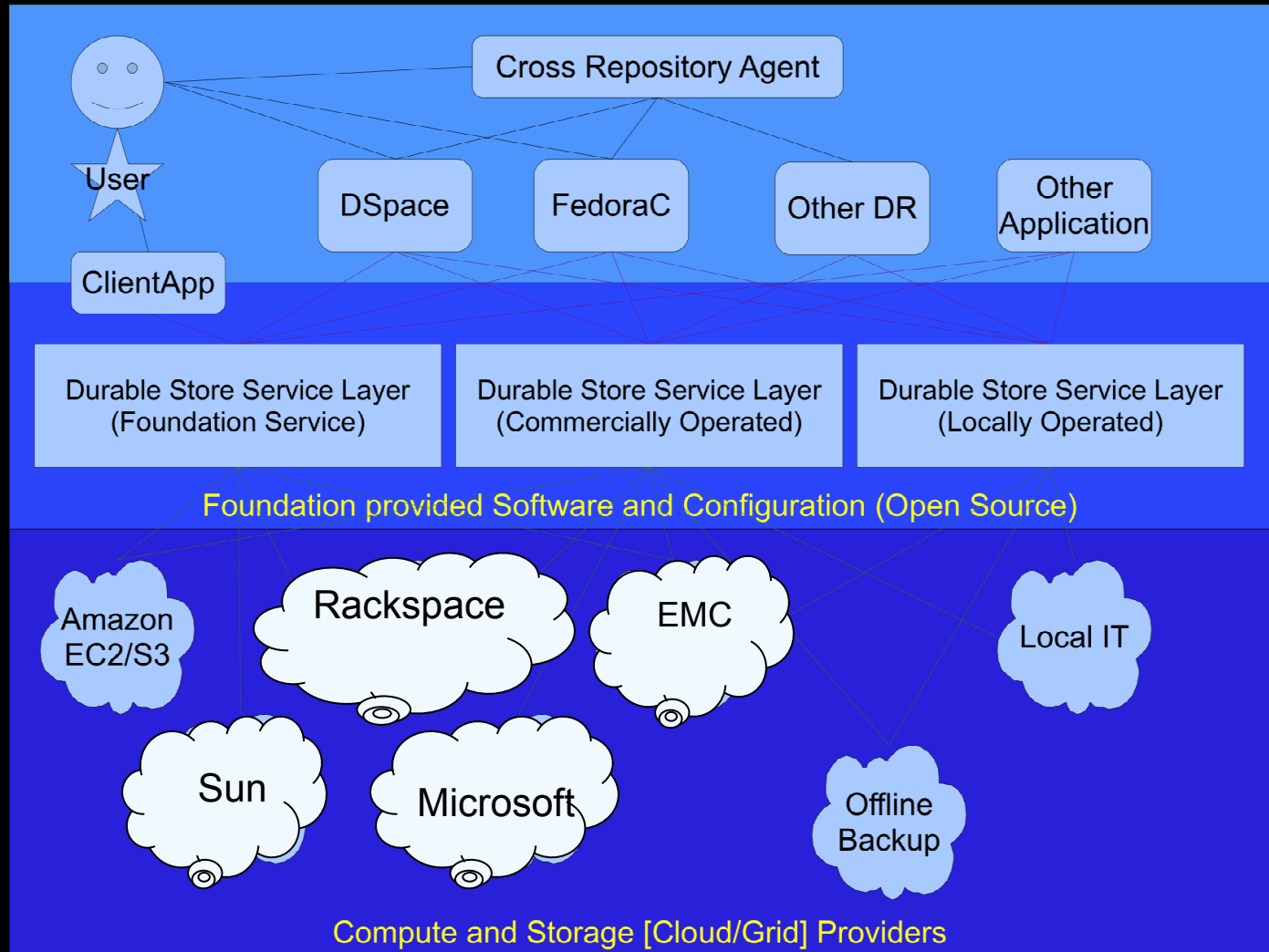
## *Trust and durability in the cloud*

DuraCloud is a platform aimed at supporting libraries, universities, and other cultural heritage organizations that wish to provide perpetual access to their digital content. The service replicates and distributes content across multiple cloud providers and enables the deployment of services to support:

- \* access
- \* preservation
- \* re-use

# DuraCloud

*A web based service enabling management of Data in the cloud*



DuraCloud  
mediating  
web  
Service

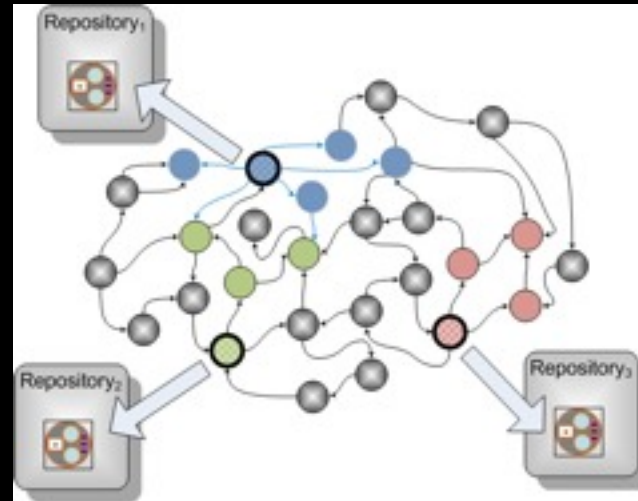
# Vision: Preservation Support



**DuraCloud: content replication, auditing, and repair**



# Vision: Shared infrastructure



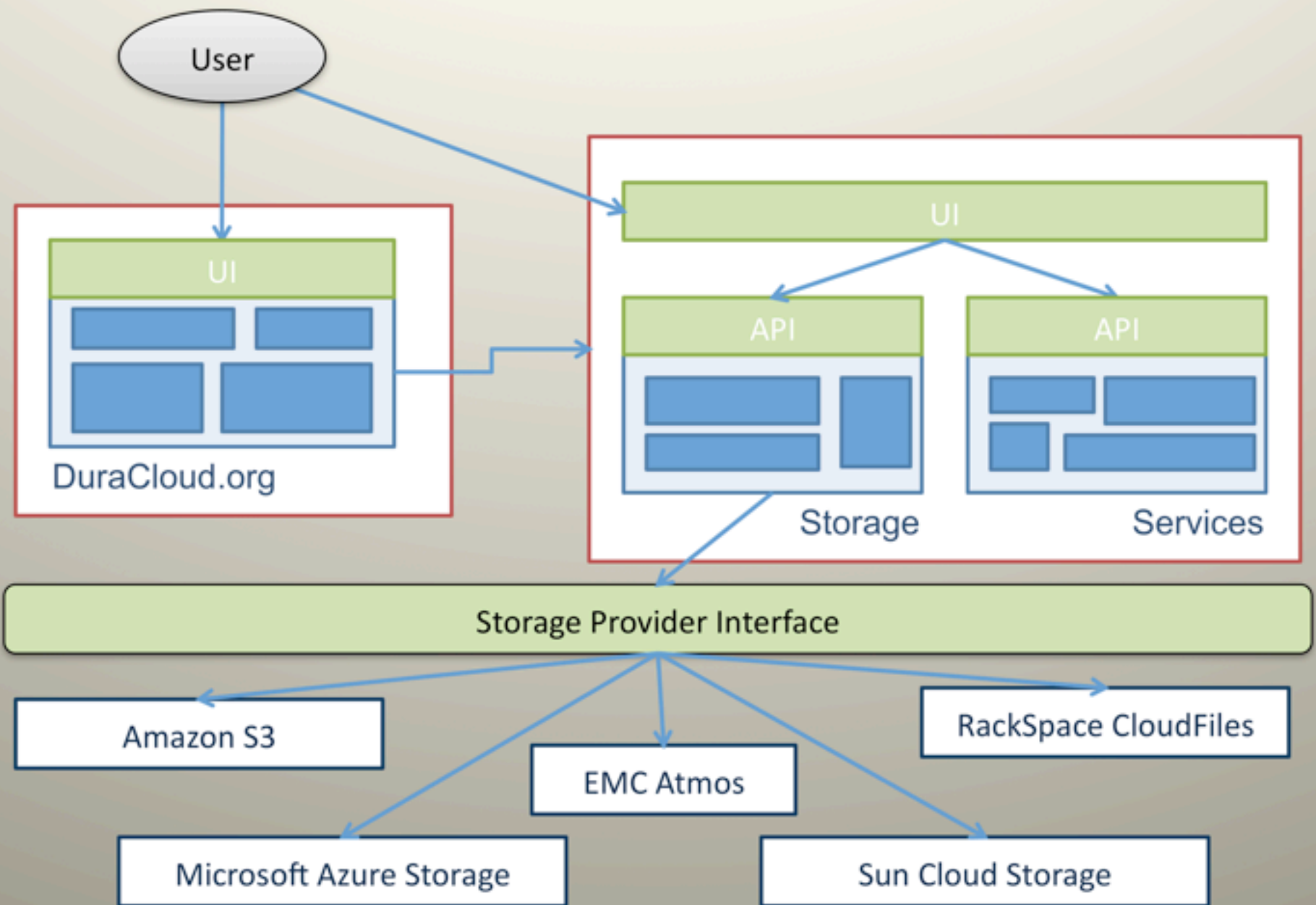
**DuraCloud: collaboration and data linking of stored objects**

# Vision: Data Analysis and Mining

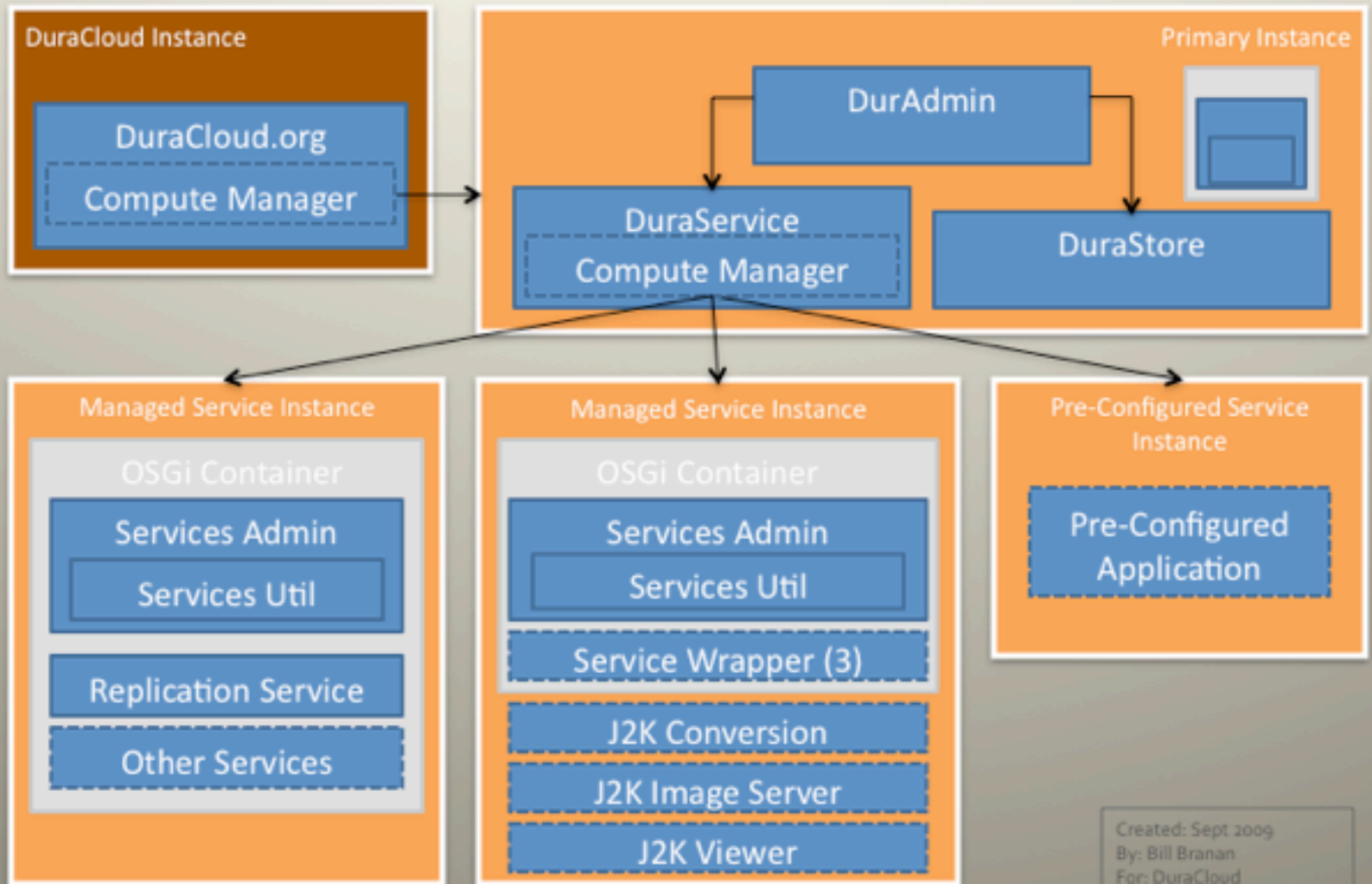


**DuraCloud: running large compute jobs on stored content**

# DuraCloud Preliminary Interaction Flow



# DuraCloud Preliminary Server Architecture



Created: Sept 2009  
By: Bill Branan  
For: DuraCloud  
<http://www.duracloud.org>

# DuraCloud

## Underlying software

- **Open core**
  - ✓ Core components available for others to build on and run
  - ✓ Open source
- **Architecture to create cloud networks**
  - ✓ Public clouds
  - ✓ Private clouds
  - ✓ University consortia
- **Also useful in research partnerships**

# Preservation Services

- ability to replicate content to multiple providers and locations*
- ability to synchronize backup with primary store or repository system*
- management ,monitoring, audit and repair through web based interface*

# software services

- Other DuraSpace-provided services on top of content stored in the cloud
  - Data mining
  - Video Streaming
  - Format transformation
  - Repository hosting
  - discovery



Enable others to build and deploy services and apps in DuraCloud environment



**DuraCloud: run your application as a service on content**

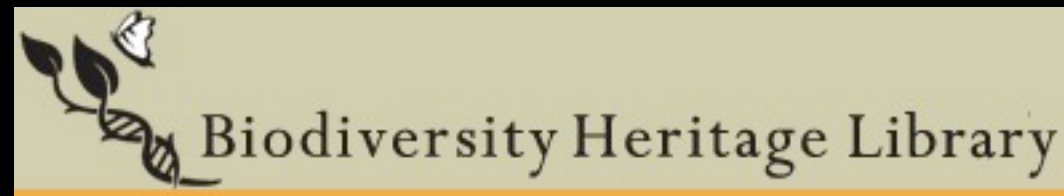


# Partners and Pilots

- Selected initial cloud providers



- Selected 2 initial pilot partners



# NYPL pilot

## *Digital Gallery Collection*

Explore NYPL's new online exhibition: *Treasures of the New York Public Library*

The New York Public Library

**Digital Gallery**

Lispenard Street between Church Street and Broadway ... Source: Drugstore Photographs. Or, A Trip Along the Yanke River, 1999, by Dylan Stone. Digital ID: 518688.

NYPL Digital Gallery provides free and open access to over 700,000 images digitized from the The New York Public Library's vast collections, including illuminated manuscripts, historical maps, vintage posters, rare prints, photographs and more.

**SEARCH**  [\(More Options\)](#)

**BROWSE**

- Subjects
- Names
- Library Divisions
- My Selections
- Arts & Literature
- Cities & Buildings
- Culture & Society
- History & Geography
- Industry & Technology
- Nature & Science
- Printing & Graphics

**DIGITAL GALLERY PICKS**

-back up copy 700k images (50 TB data)

-transformation from Tiff to JPEG 2000

-run image server in cloud

-Push JPEG 2000 back into Fedora Repository

# BHL pilot

## BioDiversity Heritage Library

**Biodiversity Heritage Library**  
News and Updates from the Biodiversity Heritage Library  
available online at <http://www.biodiversitylibrary.org>

Over 250 years of biodiversity literature to support the work of scientists, researchers, students and you.

**BHL RECENT ADDITIONS**

- L'illustration horticole : v. 20 5873 (added: 09/11/2009) - 09/11/2009
- L'illustration horticole : v. 19 5872 (added: 09/11/2009) - 09/11/2009
- The Floricultural Cabinet and Florist's Magazine. VOL. 1 (added: 09/10/2009) - 09/10/2009
- Voyage de La Pérouse autour du monde / Atlas (added: 09/10/2009) - 09/10/2009

**TUESDAY, AUGUST 25, 2009**  
**Book of the Week: BHL, EOL, and Marine Life**

The dream of making biodiversity information freely available to people around the world is an ambitious goal embraced by many in the scientific community, and it represents the mission behind both the Biodiversity Heritage Library and the Encyclopedia of Life ([www.eol.com](http://www.eol.com)), of which BHL is a cornerstone institution. While the Biodiversity Heritage Library strives to digitize the world's biodiversity literature...

**ABOUT BHL**  
Twelve major natural history museum libraries, botanical libraries, and research institutions have joined to form the Biodiversity Heritage Library. The participating libraries have over two million volumes of biodiversity literature collected over 200 years to support the work of scientists, researchers, and students in their home institutions and throughout the world.

**LABELS**  
News (3)

- back up copy entire corpus (40 TB data)
- have multiple copies including Europe
- Do compute intensive data mining over corpus

# Pilot use cases

## NYPL

Replication and preservation support

Format conversion

Instant provisioning of image server

Synchronization with repository

## BHL

Replication and preservation support

- International collaborative infrastructure

Researcher platform for data mining

# Timeline

- Begin pilots(MOU's in place) – September 2009
- DuraCloud Alpha Pilot release- Oct 2009
- Pilot data loading and testing – Fall 2009
- Beta for repository community - Q1 2010
- Pilot testing with software services Q1 2010
- Cloud partner evaluations complete-Q2 2010
- Strategic cloud partnerships in place- Q2 2010
- Pricing Model determined-Q2 2010
- Report pilot results – Q2 2010
- Launch production service Q3 2010

# Critical success factors

Ease of use- simplicity

Trusted partner for end user

Cost effective

Scalable/Flexible

Can establish key partnerships with service providers

Can build community of developers and users

# Thank You



For more information:

DuraSpace Organization: [http://  
duraspace.org](http://duraspace.org)

Wiki: [http://www.fedora-commons.org/  
confluence/display/duracloudpilot/](http://www.fedora-commons.org/confluence/display/duracloudpilot/)

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

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