DuraCloud Pilot Program: utilizing cloud infrastructure as an extension of your repository

Michele Kimpton
Project Director, DuraCloud

OR2010, Madrid Spain
July 7, 2010

DuraSpace not for profit

DURA SPACE

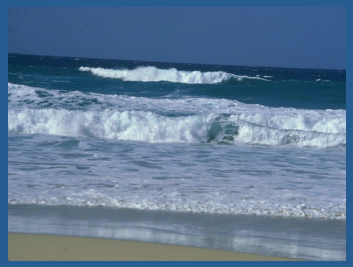
Overview

- DuraCloud platform
- Results of survey
- Pilot program
- Use cases
- Future direction

DURACLOUD™

DURA SPACE

Implications for our future work



more distributed

More collaborative

more web-oriented

more open

more interoperable

DURACLOUD

DURA SPACE

Cloud Infrastructure

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).

DURACLOUD™

DuraCloud Platform

Open technology and hosted service for utilizing cloud infrastructure for preservation support and access services

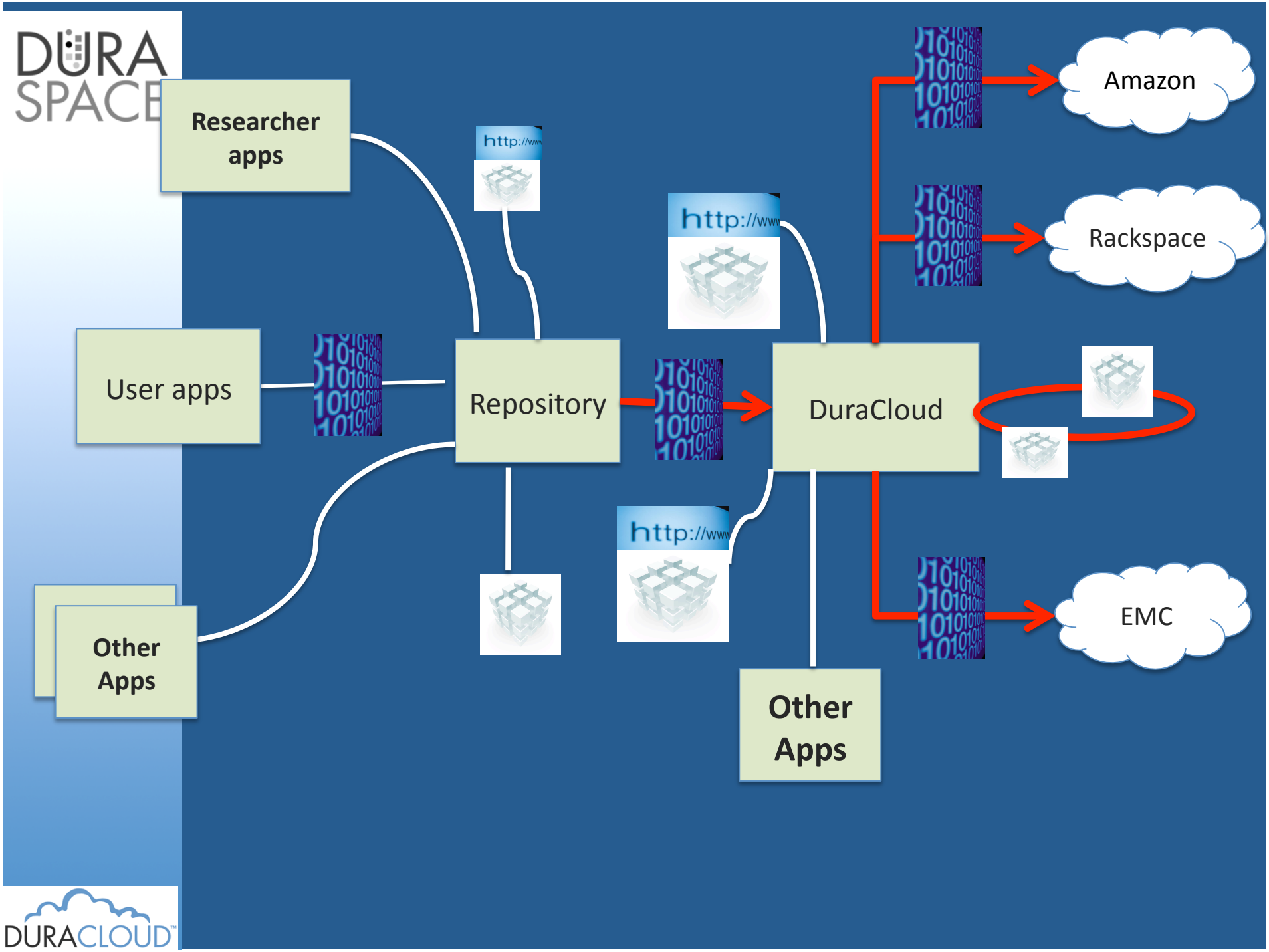
Features:

Interoperable across multiple cloud providers

Web enabled

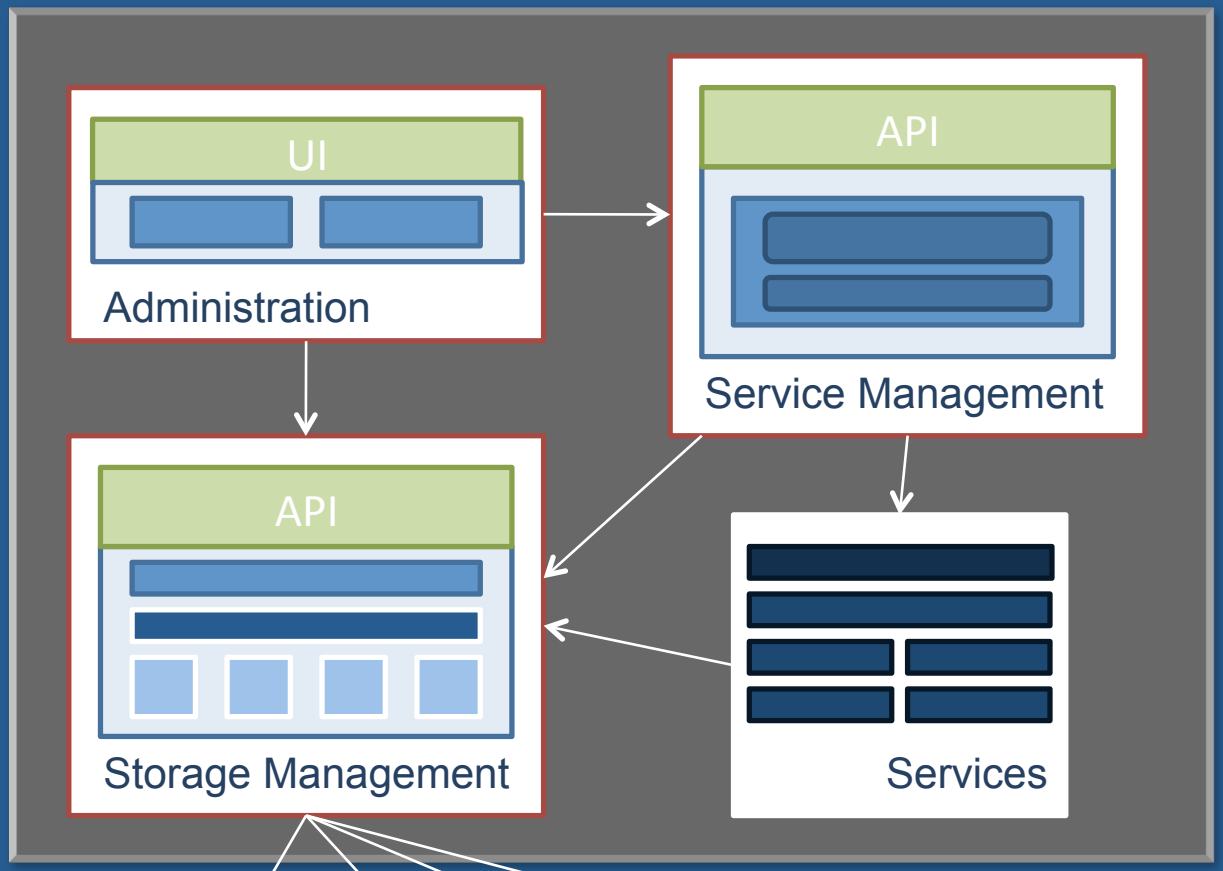
Built on highly scalable, flexible infrastructure

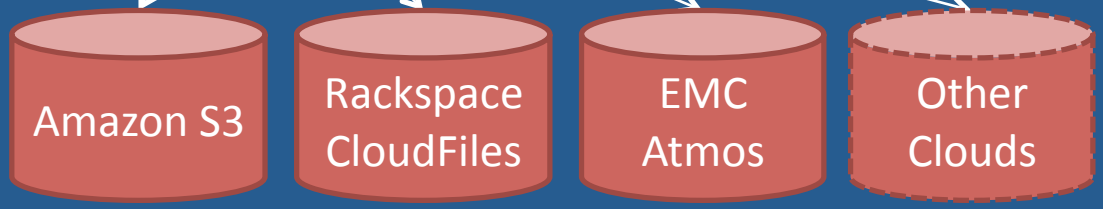
Open API's for easy integration


DURA SPACE
Researcher apps
User apps
Other Apps
Repository
DuraCloud
Amazon
Rackspace
EMC
Other Apps
DURACLOUD

DURA SPACE

DuraCloud High Level Interaction





DURACLOUD

DURA SPACE

Services and Capabilities

Replication

Image Viewing

Image Transformation

Media Streaming

Bit Integrity Checking

General Compute Services

…more on roadmap

DURACLOUD

Key Advantages Cloud

completed 1/22/2010

145 participants higher ed

Most Impactful Advantages Electronic Survey	**Responses**
Scalability	79
Remote, Off Campus Storage of Digital Assets	64
Ease of Implementation	54
Flexibility	53
Don't Have to Staff Locally	39
Cost	33
Elasticity	26
Pay for Use	14
Other	5

Key Challenges Cloud

completed 1/22/2010

145 participants higher ed

Key Challenges Electronic Survey	**Responses**
Trusting Third Party to Manage Critical Assets	64
Long-term Reliability of Solution	52
Data Security	51
Performance and Bandwidth Concerns	37
Loss of Control	34
Administrative Burden of SLAs	17
Transparency of Solution	16
Concerns about Data Lock-in	16
Less Customizable	10
Other	12

Likely to use cloud services in next 12 months

Percentage of electronic survey respondents noting it is "very likely" or "likely" they will use cloud compute or cloud storage services to manage, store or provide access to digital collections in the next twelve months.

Category		Subcategory	Percentage
Non-US			47.7%
US			51.3%
US Institutions	Institution Size	Large, very large	47.2%
		Medium	68.8%
		Small, very small	42.9%
	Enrollment Profile	RU/VH	52.1%
		RU/H, DRU	50.0%
		Master's S, M and L	46.2%
		Bac and Assoc	57.1%
	Public/ Private	Public	46.9%
		Private	59.3%

Institutional needs: managing digital collections

Service Area	Importance	Extent Need is Met	Difference	*Likelihood to Use Cloud Services*
Remote secondary storage of digital collections	3.54	2.60	0.94	*3.09*
Preservation support	3.35	2.17	1.18	*2.88*
Intra-institution shared collections	3.11	2.47	0.64	*2.69*
Inter-institution shared collections	2.72	2.07	0.65	*2.67*
Compute services	2.80	2.25	0.55	*2.54*
Online primary storage	*3.51*	*2.97*	*0.53*	*2.29*

1=low, 4=high

DURA SPACE

Purpose of Pilot Program

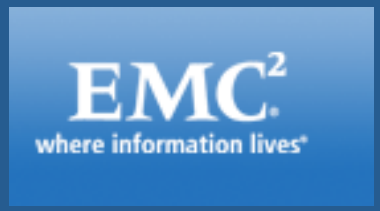
- Engage with users whom have concrete use cases and want to test the software
- Real data at scale
- Uncover the obstacles
- Engage community response and assessment

DURACLOUD™

DURA SPACE

Partners and Pilots

- Selected initial cloud providers



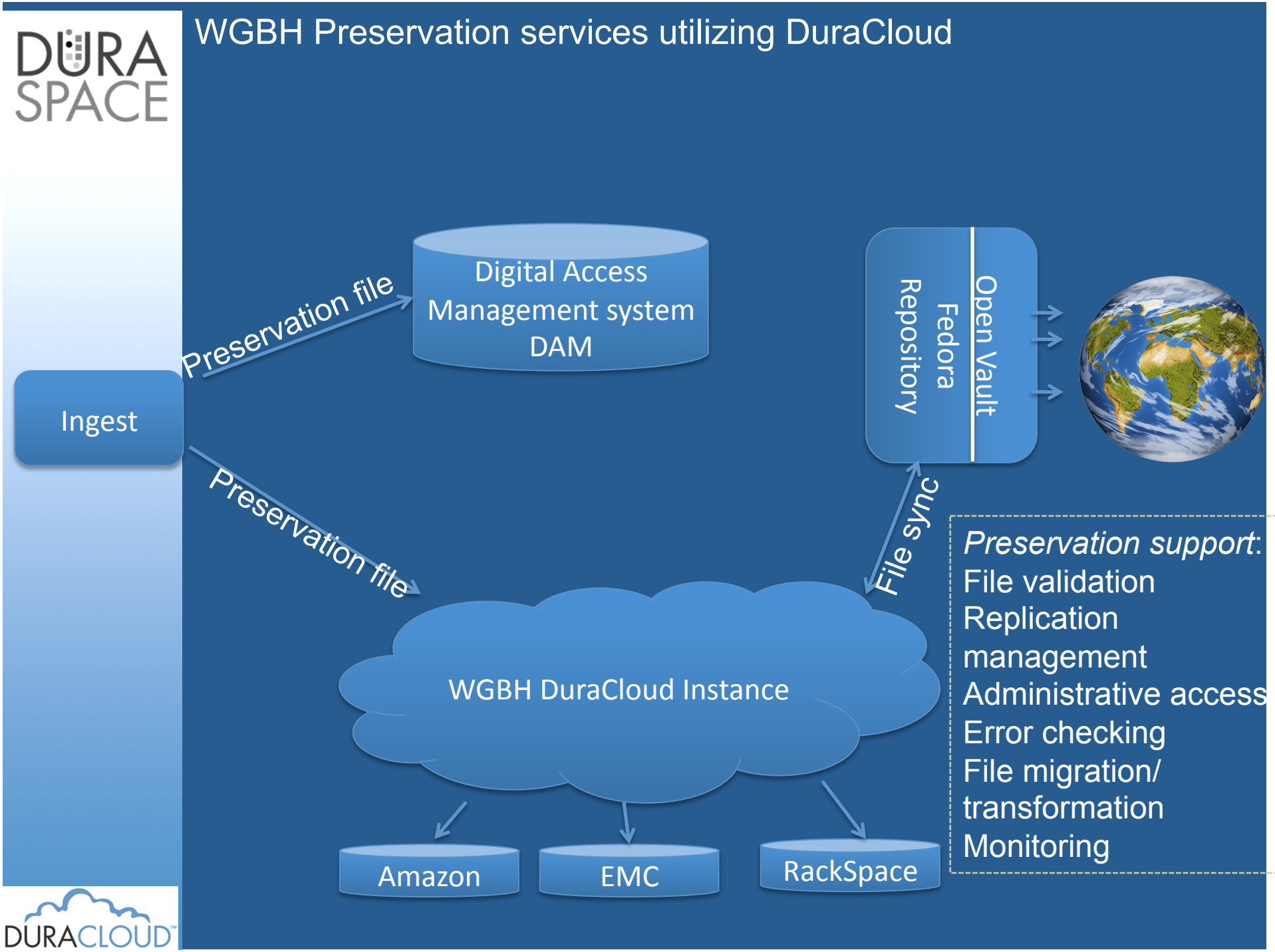
- Selected 3 initial pilot partners

DURACLOUD™

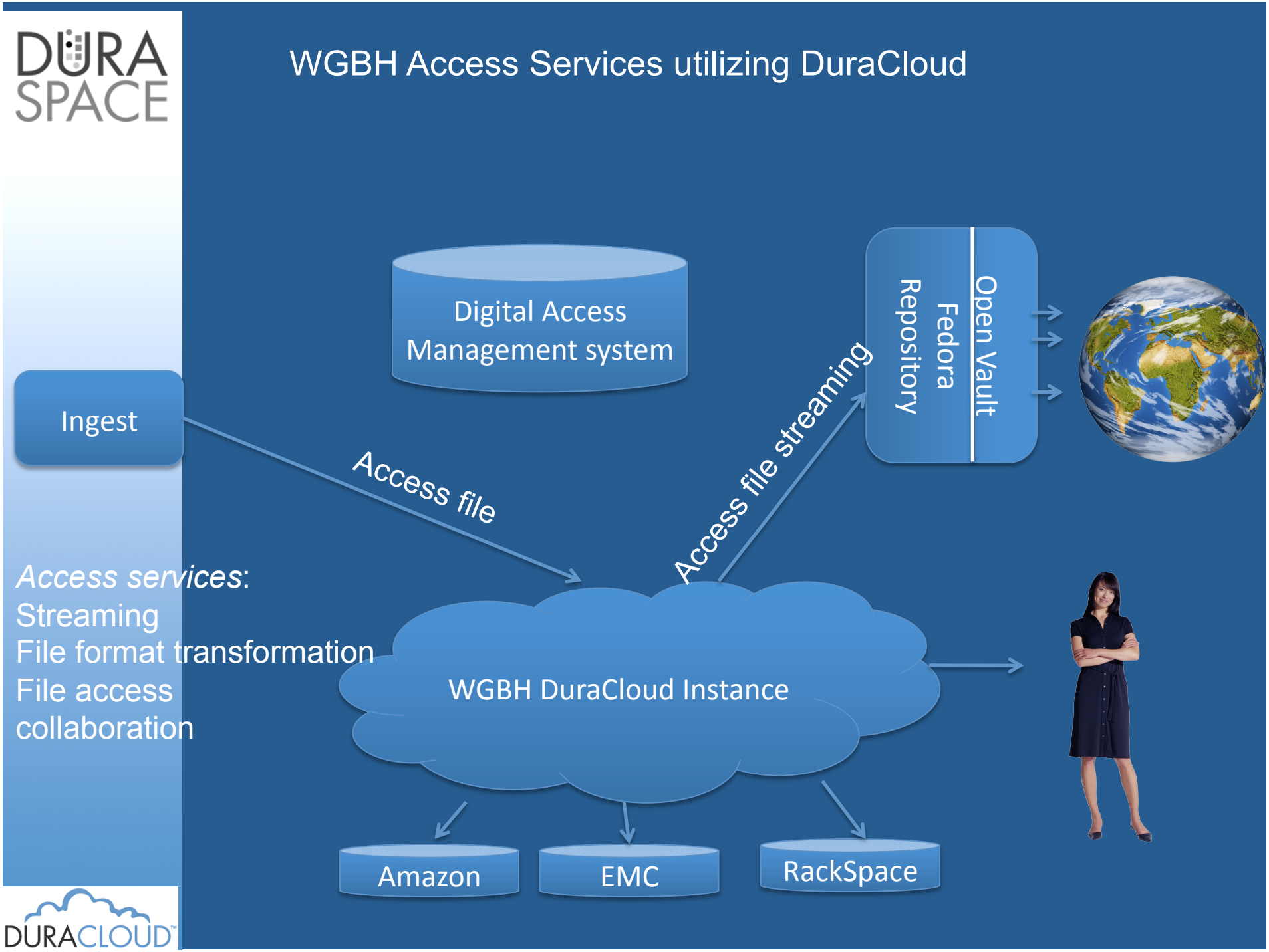
WGBH DuraCloud Pilot Goals

- Preservation
  - Uncompressed video and audio archival storage
  - Monitor and audit content
  - Replicate to multiple locations
- Access
  - Streaming audio and video
  - Video editing, transcoding
  - Researcher and third party access

WGBH Preservation services utilizing DuraCloud



WGBH Access Services utilizing DuraCloud

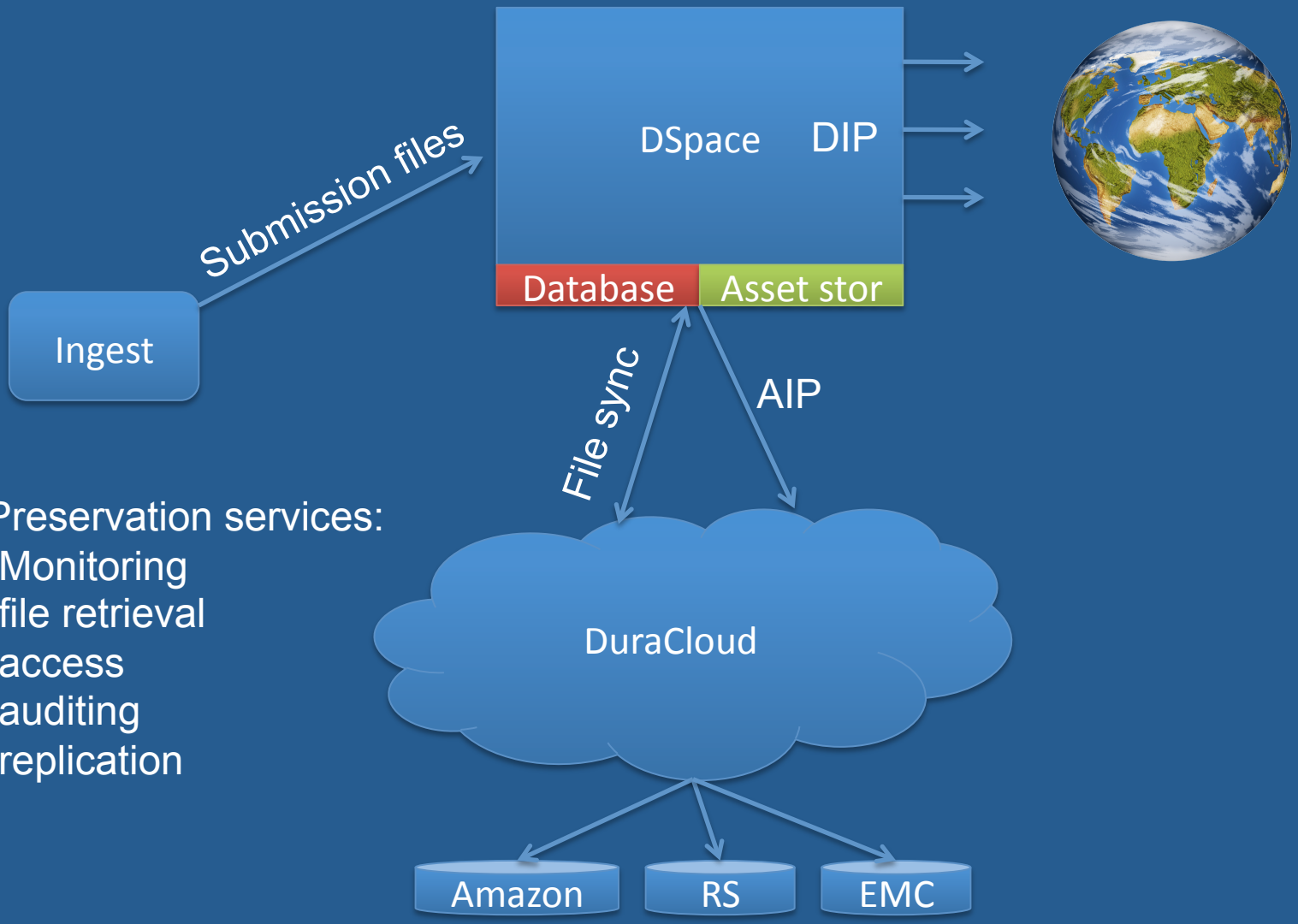


Access services:
Streaming
File format transformation
File access collaboration

DURA SPACE

MIT DuraCloud use case, preservation support

-Retrieval of lost files (admin error)
-Replacement of damaged files



Preservation services:
-Monitoring
-file retrieval
-access
-auditing
-replication

DURACLOUD

DURA SPACE

Lessons Learned

- Storage is competitive, bandwidth can be costly
- Internet Latency is high
  - Minimize transactions across the wire
  - Data should be close to compute
- Files must be less than 5 gigabytes
- Market still developing

DURACLOUD

DURA SPACE

Best "fit" use cases for Repositories

- Preservation and management
  - Second copy in cloud
  - Synchronization with primary store
  - Audit and monitoring
- Researcher, or 3rd party access
- Easy Provisioning for storage or compute

DURACLOUD™

Pilot Partners

University	Use Case	Repository
Rice U	Preservation	DSpace, meta archive
Hamilton College	Access/international collaboration	Fedora
Northwestern U	Preservation books, audio, image	Fedora
U of PEI	Image viewing/hosting	Fedora/Islandora
Cornell U	Data stream access and preservation	Fedora
ICPSR	Access and Preservation	Fedora
SUNY Buffalo	Preservation	DSpace
IUPUI	Preservation	DSpace
Rhodes College	Image Access	DSpace
North Carolina State U	Preservation	DSpace
CARL	Preservation and Services	Fedora
Orbis Cascade Alliance	Preservation and Services	DSpace
MIT	Preservation, OAIS compliance	Dspace
NYPL	Preservation and Services	Fedora
WGBH	Access and Preservation	DAM

DURA SPACE

Timeline

- Begin pilots– Oct 2009
- DuraCloud Alpha Pilot release- Oct 2009
- Initial pilots complete– July 2010
- Expanded pilot begins– Summer 2010
- Code available open source-TODAY
- Pilot testing with software services -Fall 2010
- Cloud partner evaluations complete-Winter 2010
- Report pilot results – Winter 2010
- Launch hosted service Winter 2011

DURACLOUD™

DURA SPACE

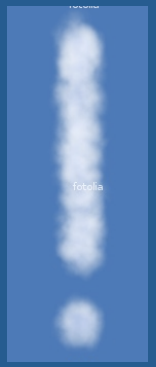
DuraCloud now available open source

- Open core
  - ✓ Open API
  - ✓ Open Source
  - ✓ Apache-style license
- Architecture to create cloud networks
  - ✓ Public clouds
  - ✓ Private clouds
  - ✓ University consortia
- Partner implementations/Integrations

DURACLOUD™

DURA SPACE

Thank You



For more information:
Come to our BOF following this session

DuraSpace organization: http://duraspace.org
Wiki: https://wiki.duraspace.org/display/duracloud/
DuraCloud project page: http://duracloud.org

DURACLOUD™

DURA SPACE

Go Spain!

DURACLOUD™