DURA SPACE

DuraCloud Pilot Program: utilizing cloud infrastructure as part of your preservation strategy

Michele Kimpton
Project Director, DuraCloud
NDIPP Partner meeting
July 21, 2010

DURACLOUD™

DuraSpace not for profit

DURA SPACE

Overview

- What is DuraCloud
- Results of survey
- Pilot program
- Use cases
- Future direction

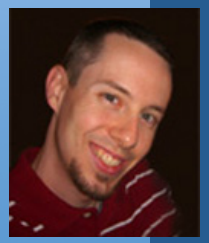
DURACLOUD™

DURA SPACE

Future Direction, Disruptive Technologies

Interoperable
Web enabled
Distributed
collaborative



DURACLOUD

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 Platform

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

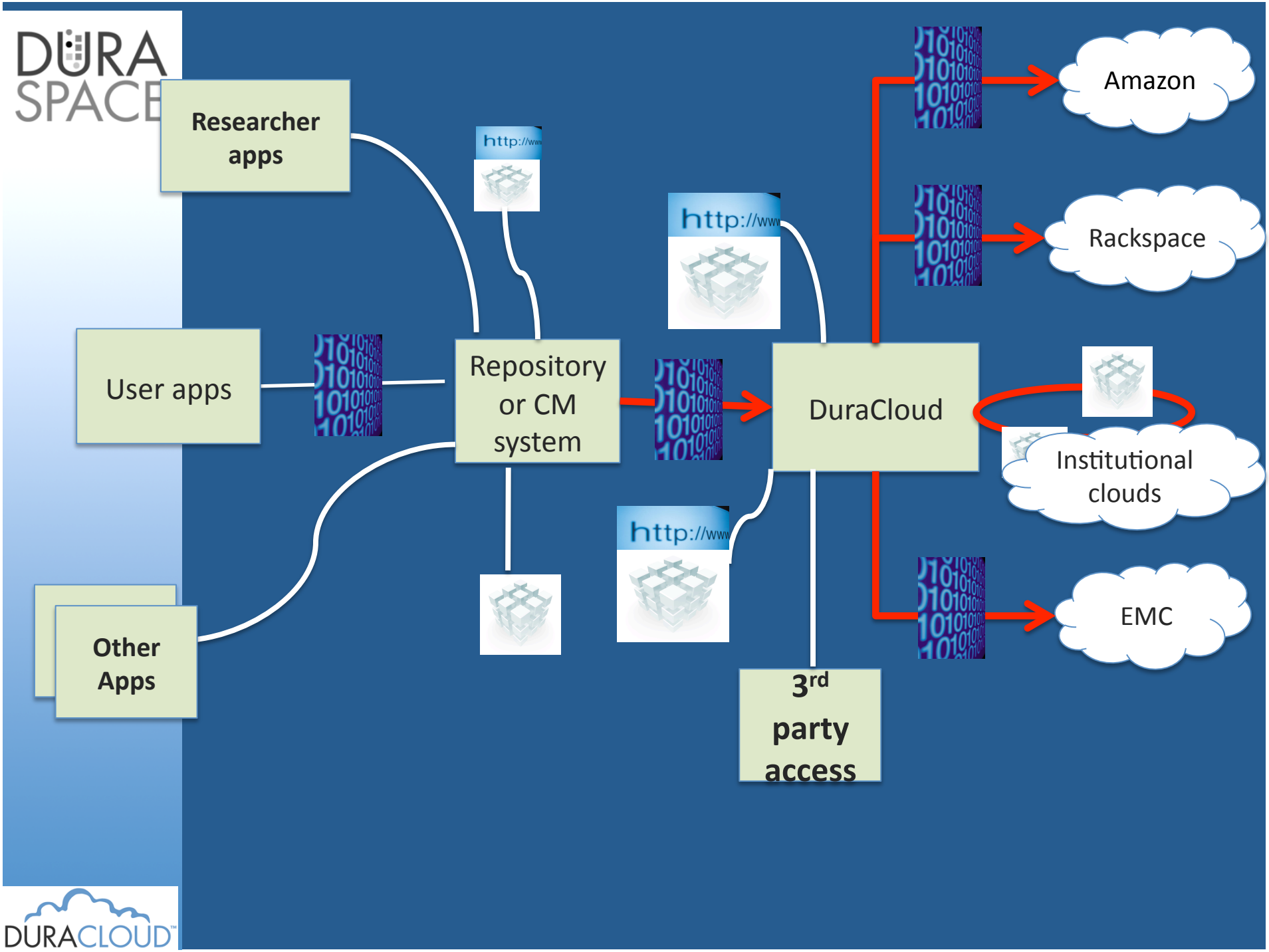
Architectural Features:

Interoperable across multiple cloud providers

Web enabled

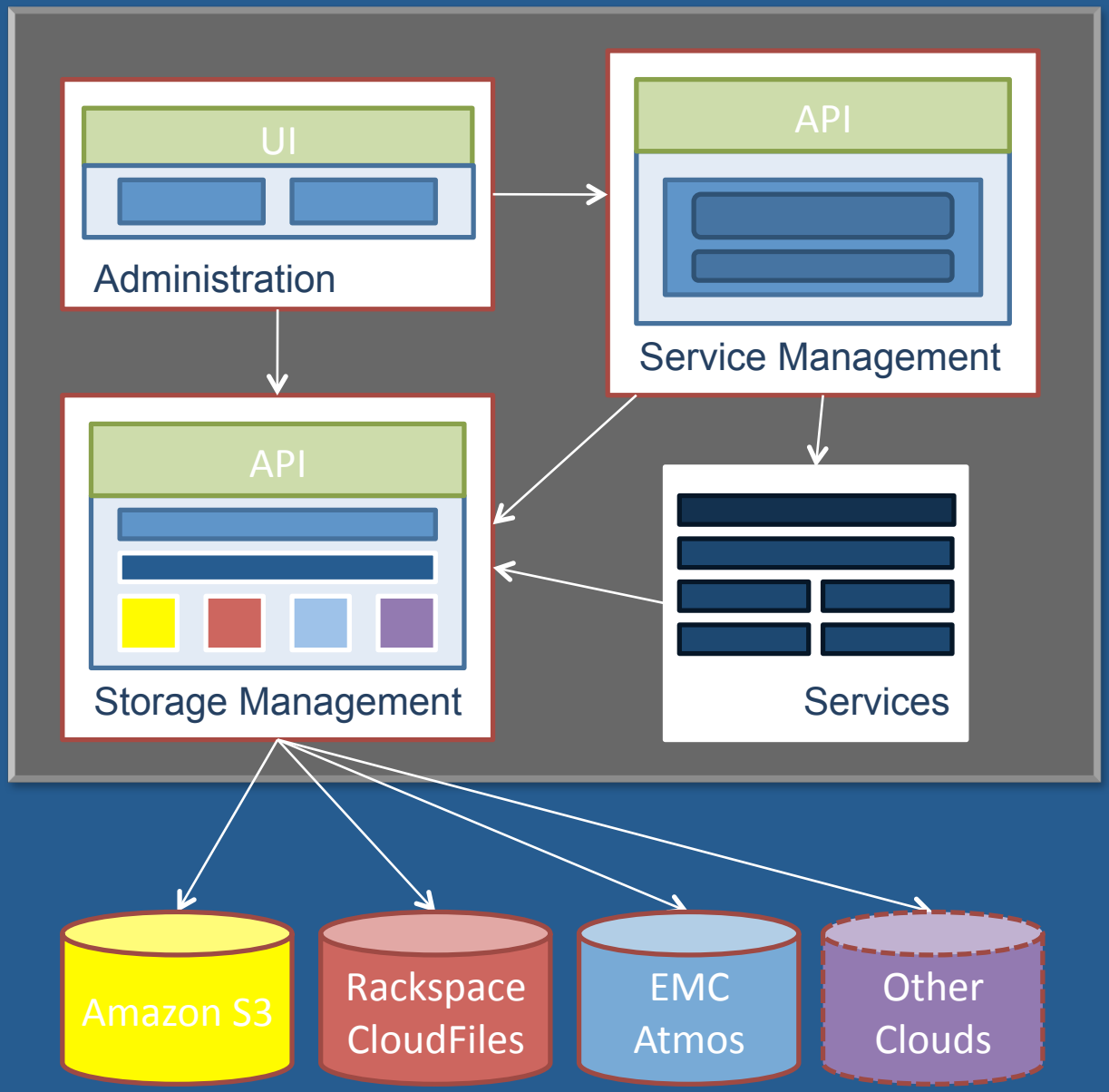
Built on highly scalable, flexible shared infrastructure

Open API's for easy integration


DURA SPACE
Researcher apps
User apps
Other Apps
Repository or CM system
DuraCloud
Amazon
Rackspace
Institutional clouds
EMC
3rd party access
DURACLOUD

DURA SPACE

DuraCloud High Level Interaction



DURACLOUD

DURA SPACE

Services and Capabilities

Replication

Image Viewing

Image Transformation

Media Streaming

Bit Integrity Checking

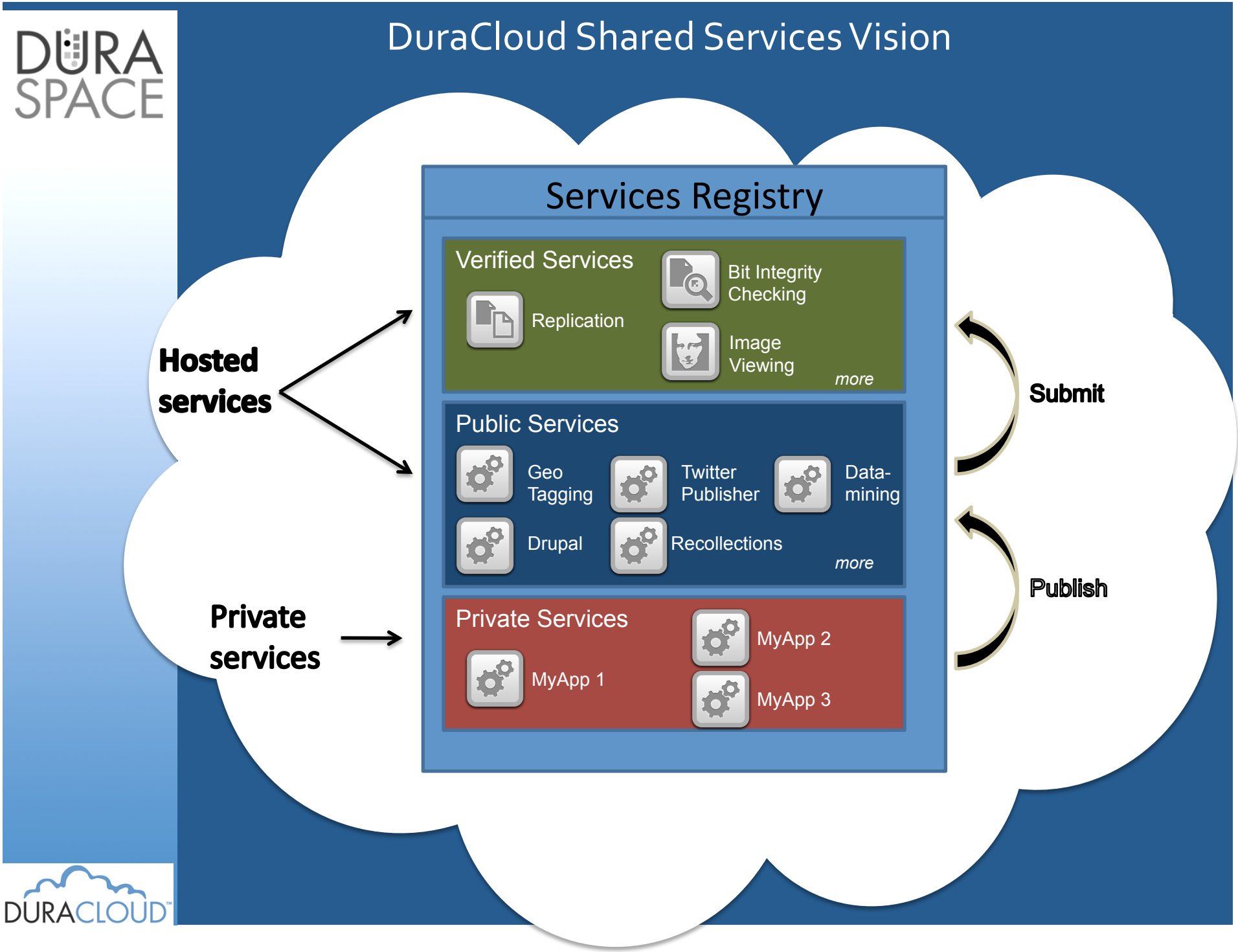
…more on roadmap

File format validation

Parallel processing

DURACLOUD

DuraCloud Shared Services Vision



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

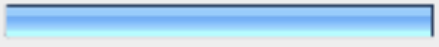
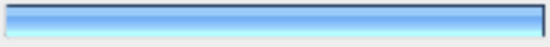
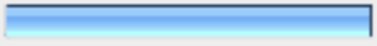
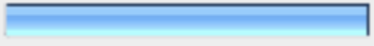
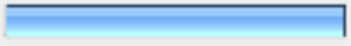
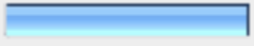
Key Challenges Electronic Survey	**Responses**
Trusting Third Party to Manage Critical Assets	64
Long-term Reliability of Solution	52
Data Security	51
Performance and Bandwidth Concerns	37
Loss of Control	34
Administrative Burden of SLAs	17
Transparency of Solution	16
Concerns about Data Lock-in	16
Less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

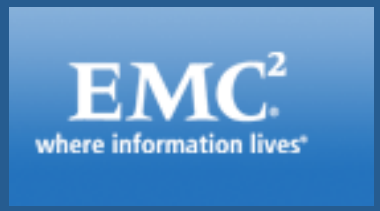
Preservation support

11. Please identify the elements of preservation support that are of most value to your institution.		Response Percent	Response Count
Replication		59.5%	91
Multiple copies managed under multiple administrations		41.2%	63
Online access		**75.2%**	**115**
File migration		51.0%	78
File validation		50.3%	77
Bit integrity checking		47.1%	72
File repair		33.3%	51
	Other (please specify)		10
	answered question		**153**
	skipped question		**71**

DURA SPACE

Partners and Pilots

- Selected initial cloud providers



- Selected 3 initial pilot partners

DURACLOUD™

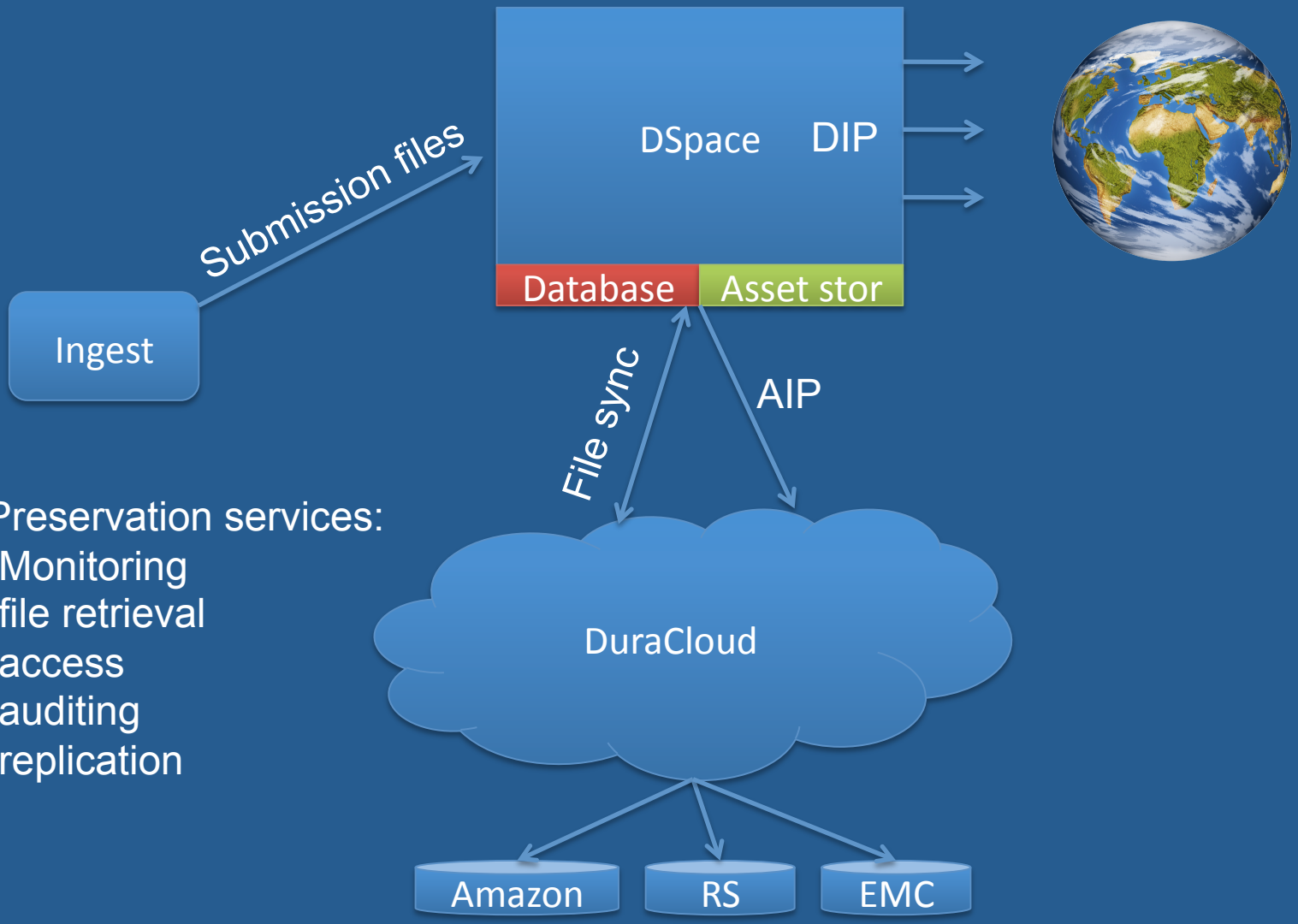
Extended Pilot Partners

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

DURA SPACE

MIT DuraCloud use case, preservation support

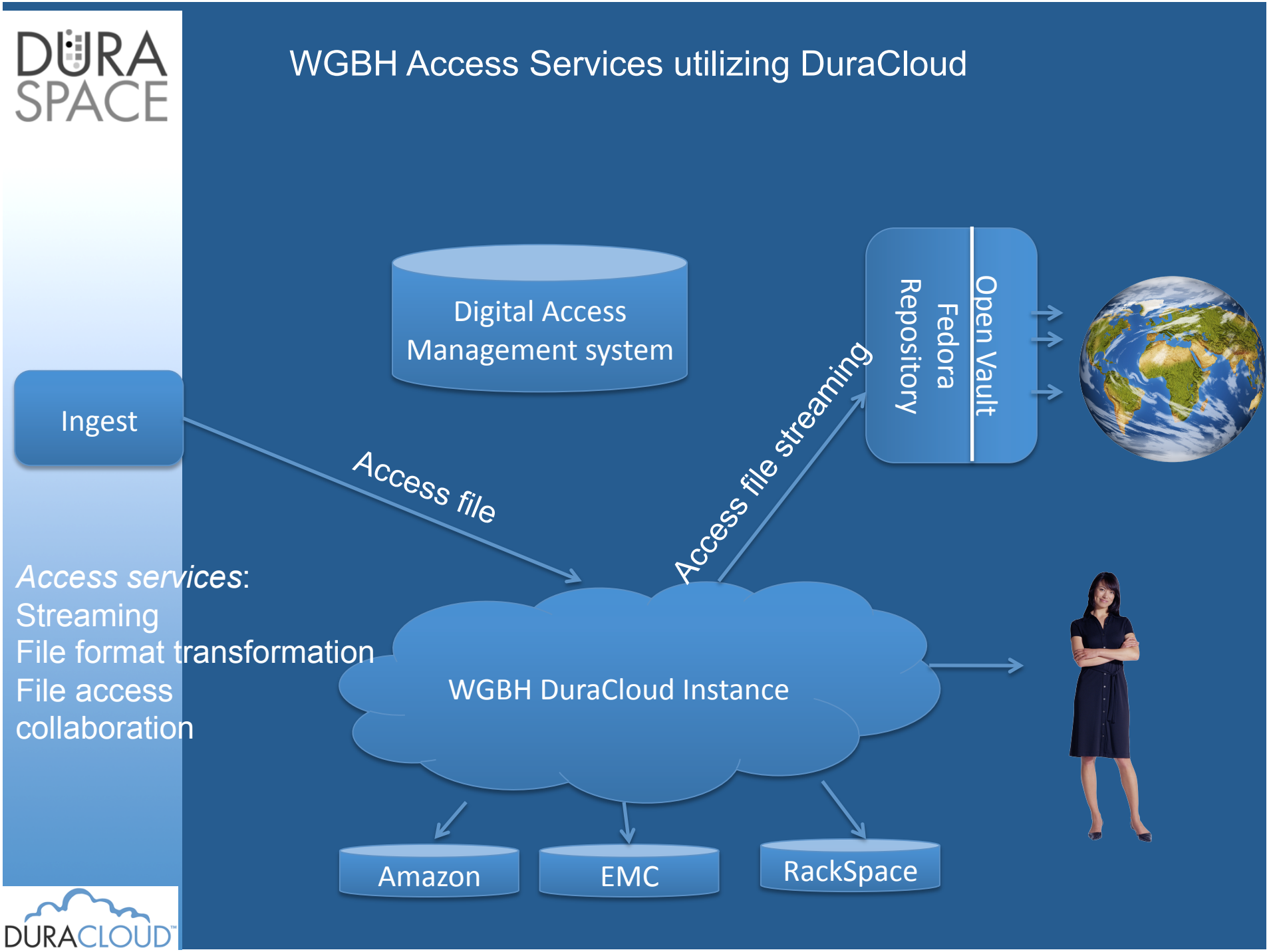
-Retrieval of lost files (admin error)
-Replacement of damaged files



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

DURACLOUD

WGBH Access Services utilizing DuraCloud



Access services:
Streaming
File format transformation
File access collaboration

DURA SPACE

Achievements during Initial Pilot

- DuraCloud integrated with 3 cloud storage providers
- Pilot partners loaded 30 TB into Duracloud
- Integrated and deployed multiple independent services
- Developed tools to overcome limitation of 5 GB file size and ease data loading

DURACLOUD™

DURA SPACE

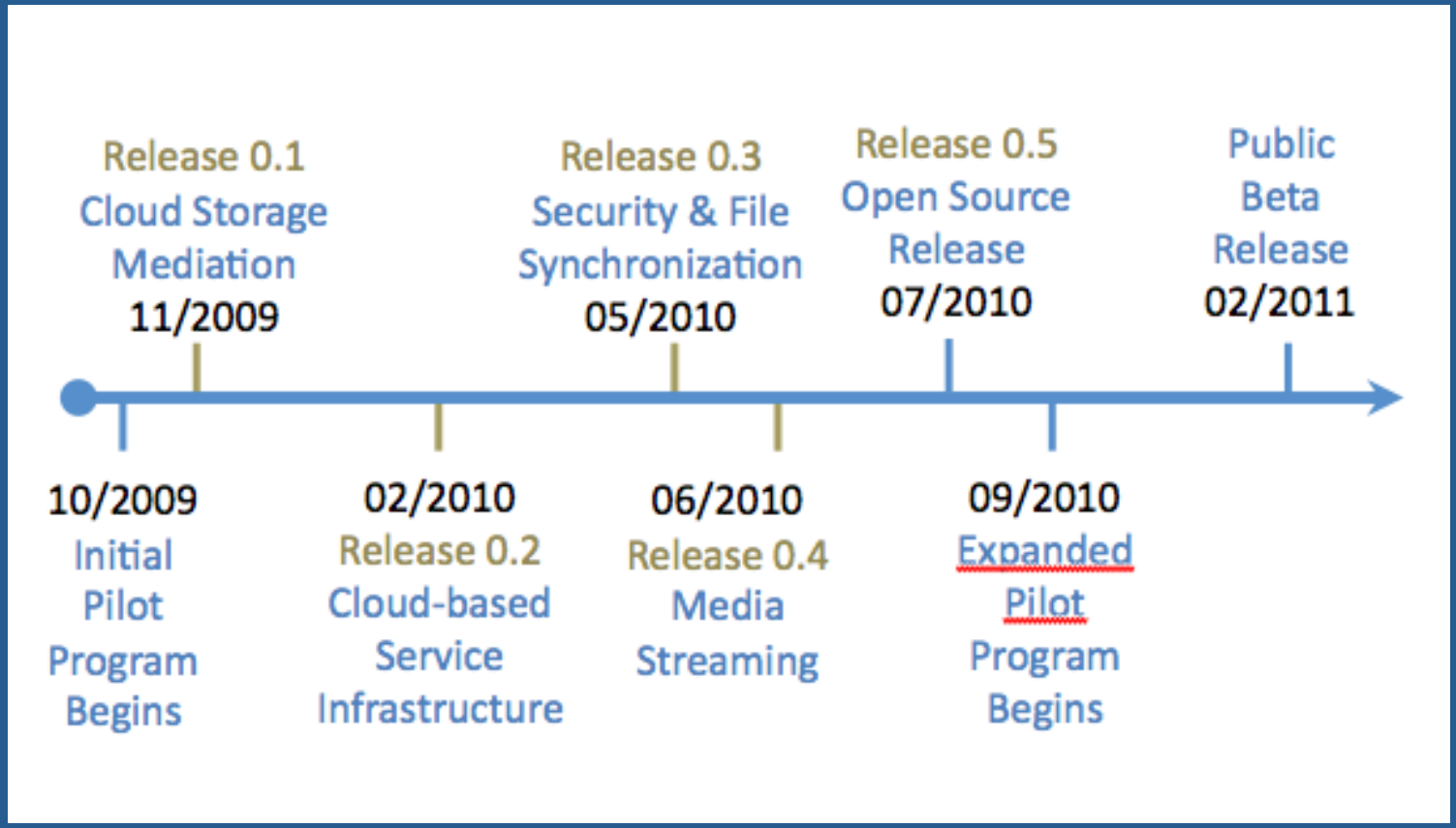
Lessons Learned

- Content transfer can be time intensive
- Internet Latency is high
  - Minimize transactions across the wire
  - Data should be close to compute
- Storage more mature than compute
- Market still developing

DURACLOUD

DURA SPACE

Key milestones



DURACLOUD

DURA SPACE

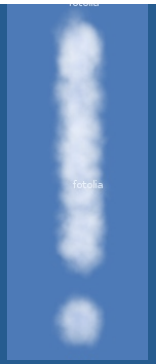
DuraCloud now available open source

- 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 Pilot Partner panel at 2:45 pm today

DuraSpace organization: www.duraspace.org
Wiki: wiki.duraspace.org/display/duracloud/
DuraCloud project page: duracloud.org
DuraCloud demonstration: https://demo.duracloud.org/durastore/movie-archive/duracloud-movie-large-take-5.m4v
DuraCloud open source: wiki.duraspace.org/display/duracloud/DuraCloud

DURACLOUD