

# DuraCloud pilot program

Michele Kimpton, CEO DuraSpace

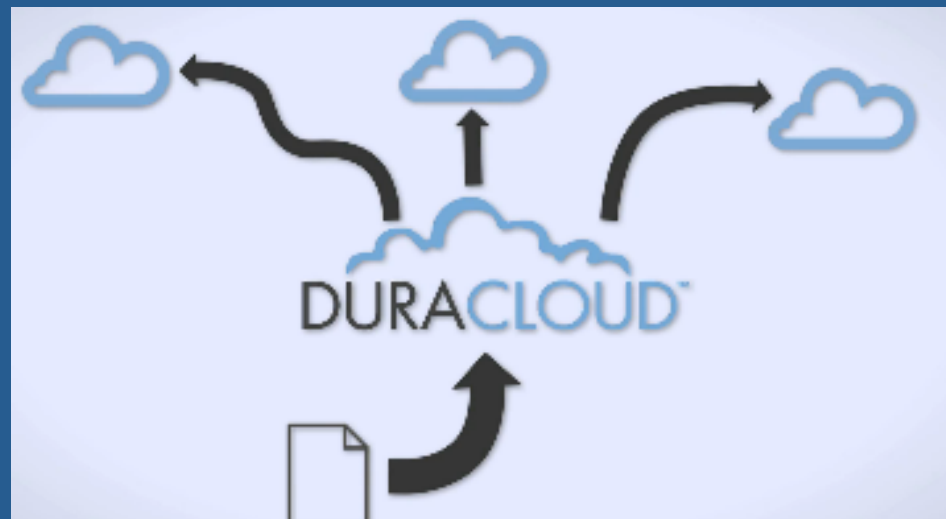
Richard Rodger,  
Dept Head Software development,  
M.I.T. Libraries

Claire Stewart  
Dept Head Digital Collections,  
Northwestern University University

OR2011 June, 2011

# What is DuraCloud

Open technology and managed service that utilizes cloud infrastructure for preservation support and access services



*Upload once and make multiple copies to various storage providers*

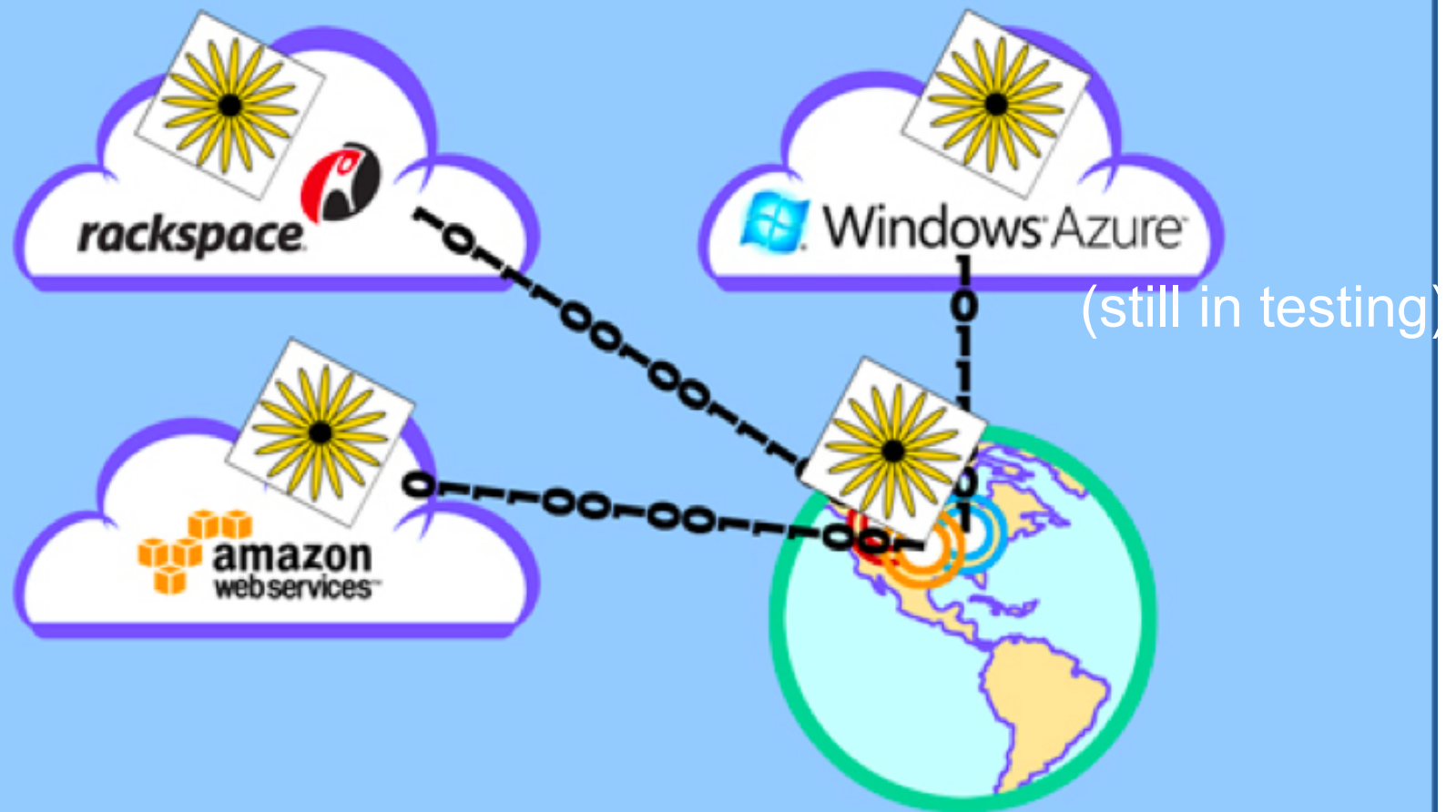
*Access anywhere from any internet connected device*



## Features:

- ✧ **Simple:** upload once and make multiple copies to various storage providers
- ✧ **Hosted service:** web-enabled, no client software needed
- ✧ **Fully managed solution:** one unified interface across multiple providers
- ✧ **Flexible:** built on highly scalable cloud platform
- ✧ **Turnkey services:** preservation support and media access
- ✧ **Open source platform**

# Underlying Providers



# Services available

## Archiving and Preservation



**Online backup  
and replication**



**Bit Integrity  
Checking**



**Duplicate on  
Demand**

## Multimedia Access



**Advanced  
Image Viewing**



**Media Streaming**



**Image  
Transformation**

## Interest for Repositories

- Take advantage of cloud at less risk
- Simple preservation support
- Take advantage of the other services

# Goals for Project

- Community driven
- Open Source and Transparent
- Easy pathway to the cloud



- Produce solutions not just technology
- Mitigate risk of using commercial cloud
- Simple and cost effective

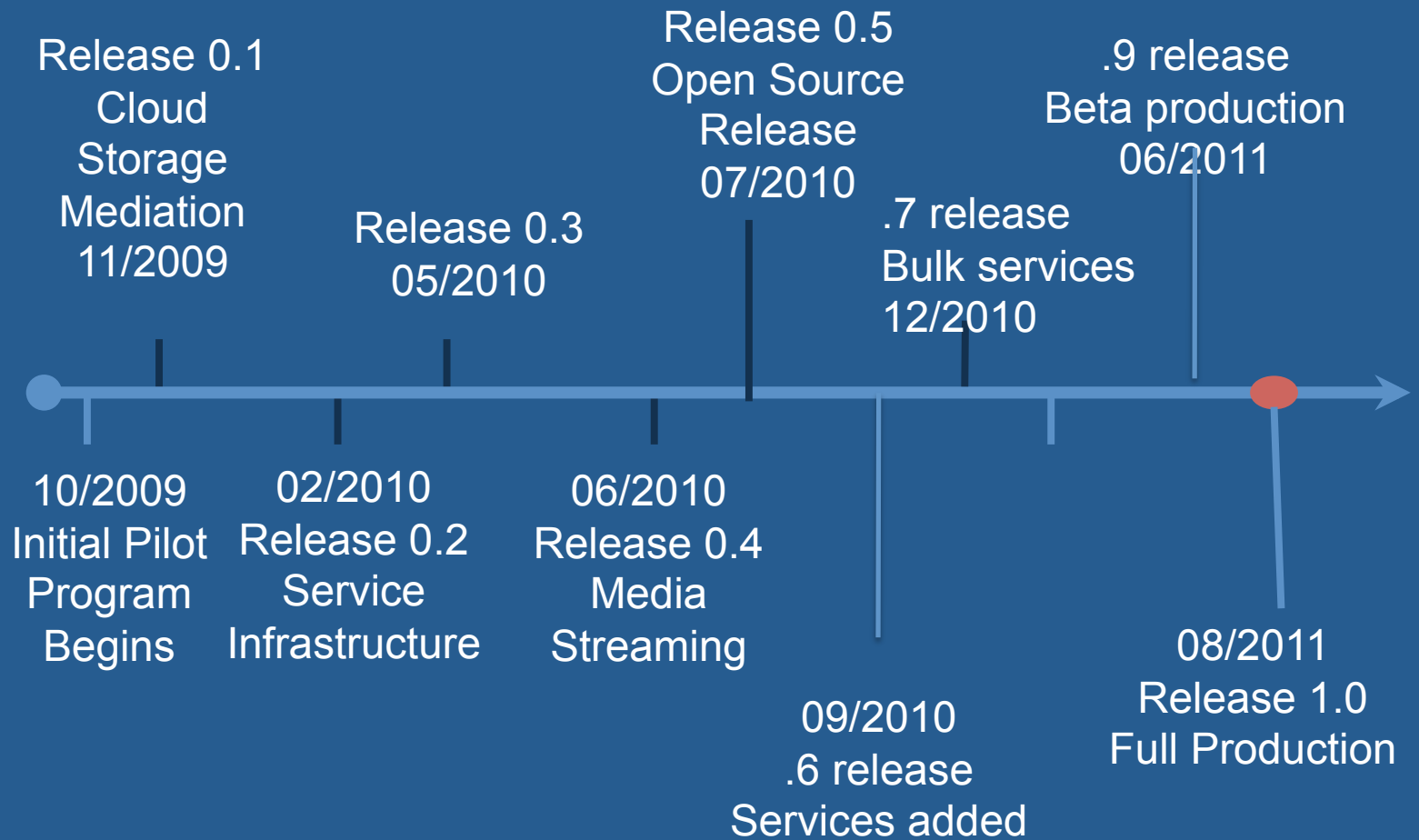
# Goals of Pilot



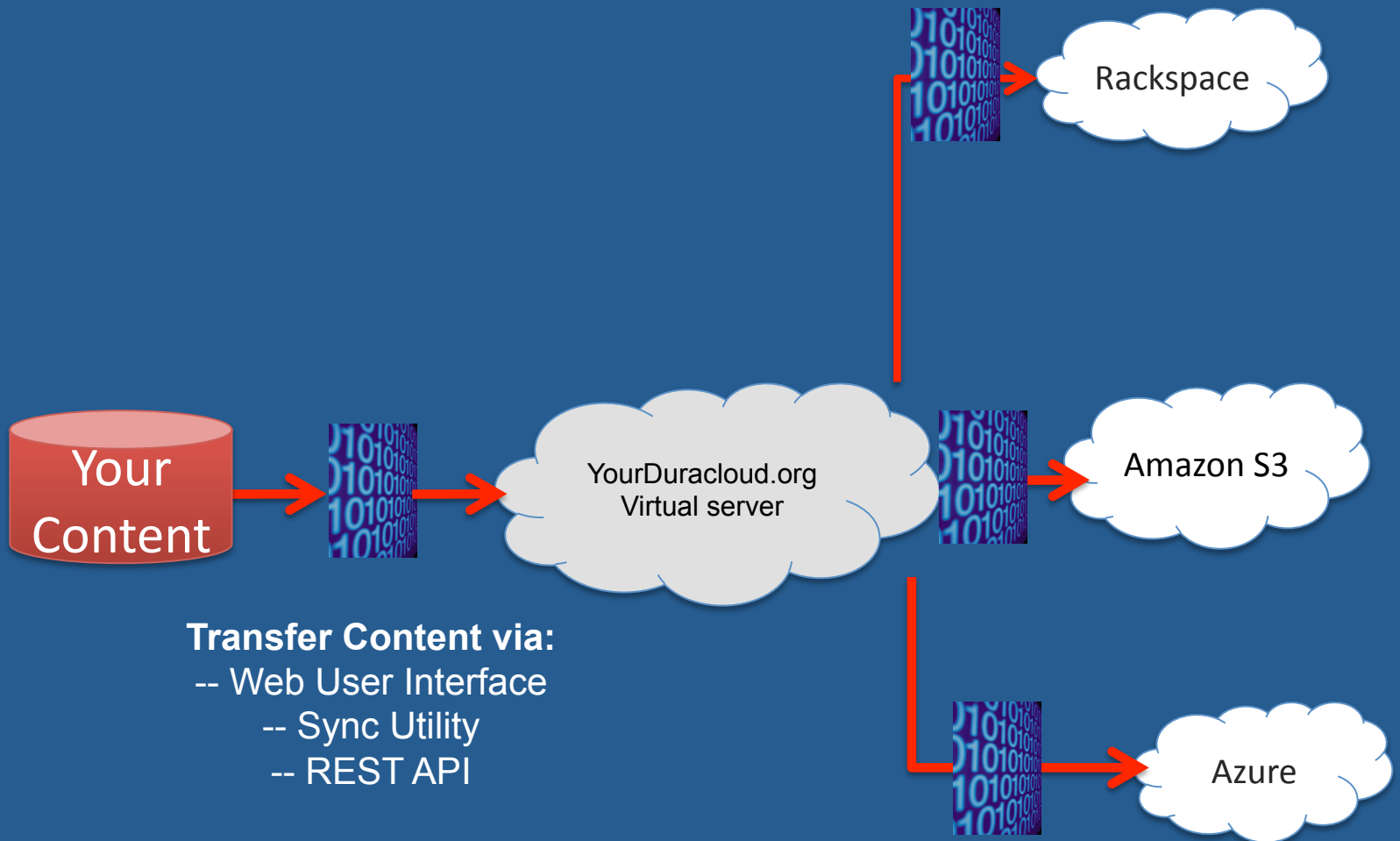
- ✓ Use Cases
- ✓ Feedback
- ✓ Validation
- ✓ Testing
- ✓ Issues



# DuraCloud Timeline



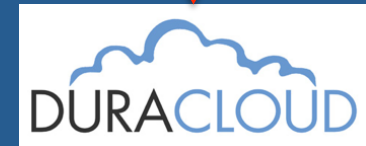
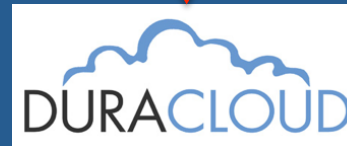
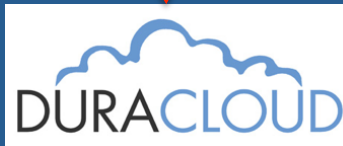
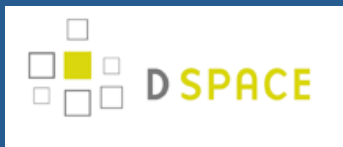
# How DuraCloud Works



# Pilot Partners

University	Use Case	Repository
Rice U	Digital archiving	DSpace, MetaArchive
Hamilton College	Access, international collaboration	none
Northwestern U	Preservation books, audio, image	Fedora
U of PEI	Digital archiving	Fedora/Islandora
ICPSR	Multiple online backups	Fedora
IUPUI	Dark online archive	DSpace, CONTENTdm
Rhodes College	Image Access	DSpace
North Carolina State U	Digital archiving	DSpace
Colorado Alliance	Preservation and Services	Fedora
Chronopolis	Integration with Grid	IRODS
MIT	Online replication and easy file restore	DSpace
Columbia	Online archiving	IA, Fedora
WGBH	Media access and archiving	DAM

# DuraCloud Integrations tested



## What we learned about the cloud

- Durability- No loss of data over 2 year period, 15+ accounts, 15+ TB
- Data availability-Internet Latency is a factor
- Transfer rates- highly dependant on user premise
- Monetary risks- runaway jobs, user attacks
- Crude access controls-S3 and cloudfront

## Key Findings from Pilots

- Importance of “Archival Information Package” for Integration with existing systems
- Many of the pilots preferred fixed allocations vs. pay as you go, few could use the cloud directly
- Biggest value to users was simplicity
- Biggest need was increased reporting and management tools



# DuraCloud @ MIT Libraries

Pilot Program Experience

Richard Rodgers

Software Development and analysis, MIT

Open Repositories 2011

# MIT Libraries Use Case

- Core: replication for preservation in our production IRs
- System administration practices only address HW or admin failures – repo errors unsecured
- Service must be both automatic yet visible
- Geared towards DSpace and system admins
- Must be cost-effective, user friendly, etc



# DuraCloud Value Proposition

- Provides convenient geo-distributed copy management
- Multi-vendor model buys abstraction and lock-in prevention
- Tools and APIs for DSpace integration
- High-bandwidth access to developers
- Platform for preservation services
- Institution-friendly service terms

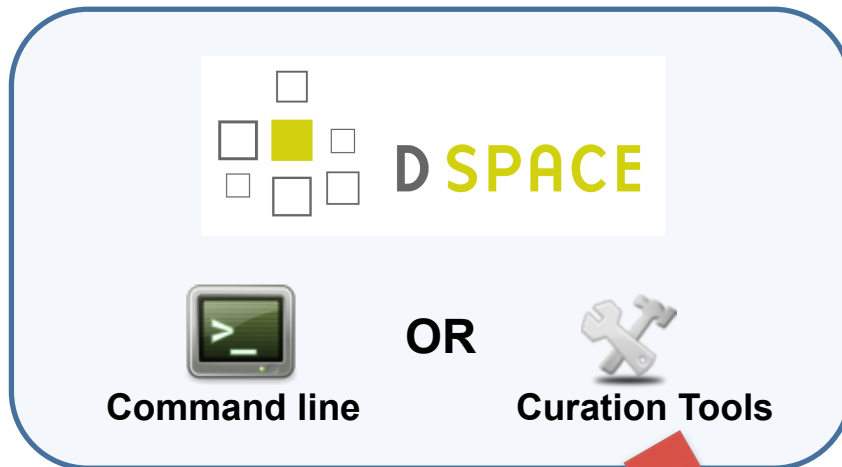
# Pilot Challenges/Solutions

- Need content containers: ALPs over loose files
- Repository managers involvement: admin UI integration, in addition to batch tools
- Bandwidth = cost control: checksum verification
- Cost visibility: activity 'dashboard'
- Service visibility: 'ad hoc' auditing

# Pilot Challenges/Solutions

- Not all replication operations interactive: queueing system
- Automatic handling of content updates: change listeners
- Deletions require special handling

# DuraCloud Integration (1.8.x)



## “Replication Task Suite”:

- Suite of Curation Tasks
- One step Sync process
- Via UI or CLI

Package for each  
Community,  
Collection & Item

Local Temp Folder  
(Cache)



# Future Activity

- Based on success of pilot, proceeding to production in 2011
- Has triggered an evaluation of preservation policy and practices – formats, extent
- Looking into specific needs and requirements around video content

# Thank you!

- Web site: [Duracloud.org](http://Duracloud.org)
- Demos at the DuraSpace Table during break
- Contact us!
  - Michele Kimpton: [mkimpton@duraspace.org](mailto:mkimpton@duraspace.org)
  - Carissa Smith: [csmith@duraspace.org](mailto:csmith@duraspace.org)

***Sign up at [DuraCloud.org](http://DuraCloud.org) for a trial account***