



# DURACLOUD™

## Preserving the scholarly record

Michele Kimpton  
CEO, DuraSpace

Stu Baker  
Northwestern University

Educause Conference  
October 19, 2011

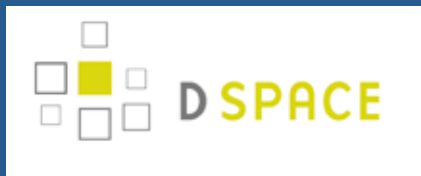


# Who are we?



# Mission

*We are committed to providing open source technologies and services that promote durable, persistent access to the scholarly record.*



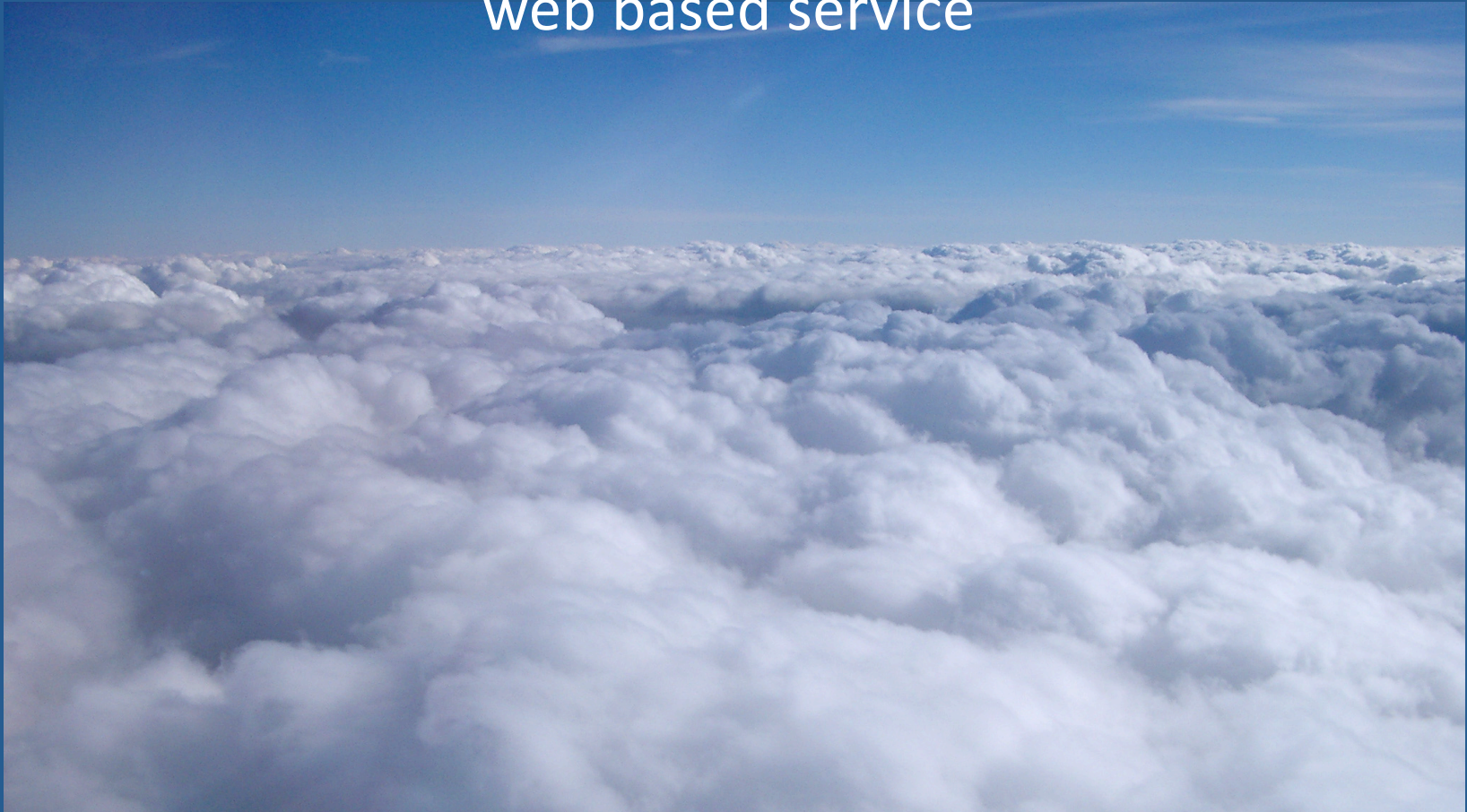
# Preservation challenges

- Ability to readily provision online storage (ideally in another geographic area, another administration)
- Synchronize content across storage systems
- Audit integrity of content
- Technical resources
- Internal Policies



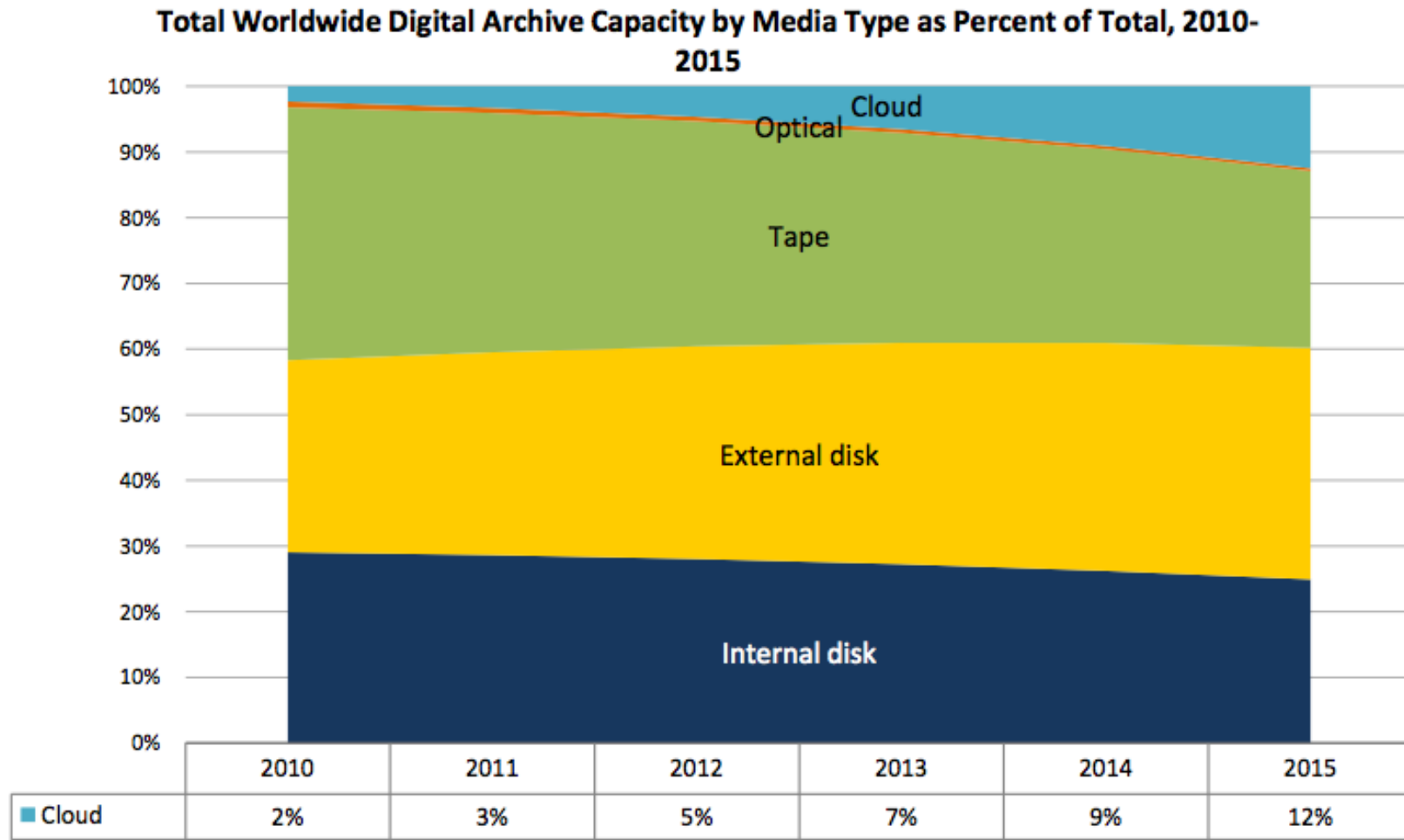
# Why cloud?

Massively scalable compute and storage offered as a web based service



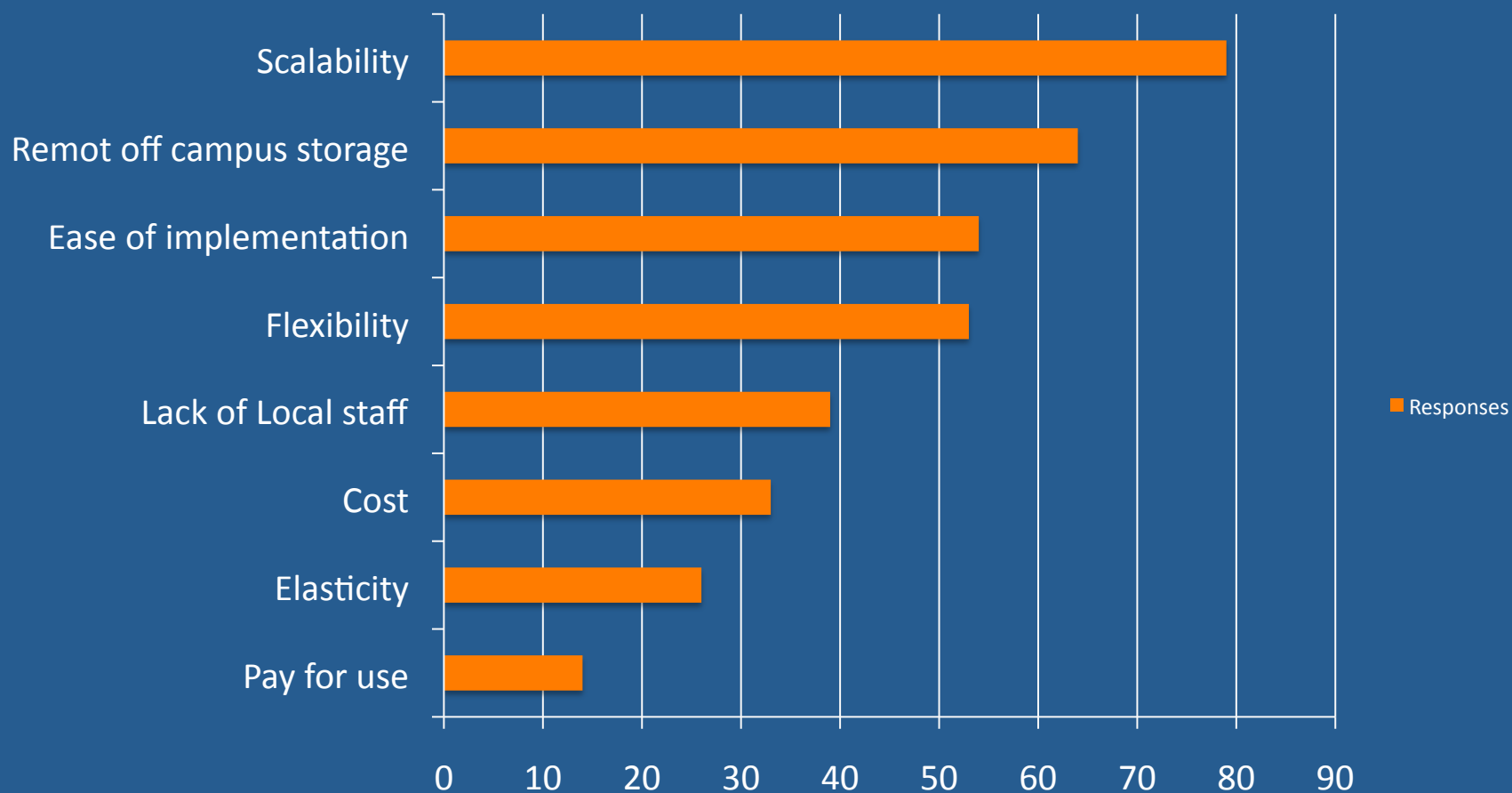
# Digital archiving by media type

Figure 2. Total Worldwide Digital Archive Capacity, by Media Type as Percent of Total, 2010-2015



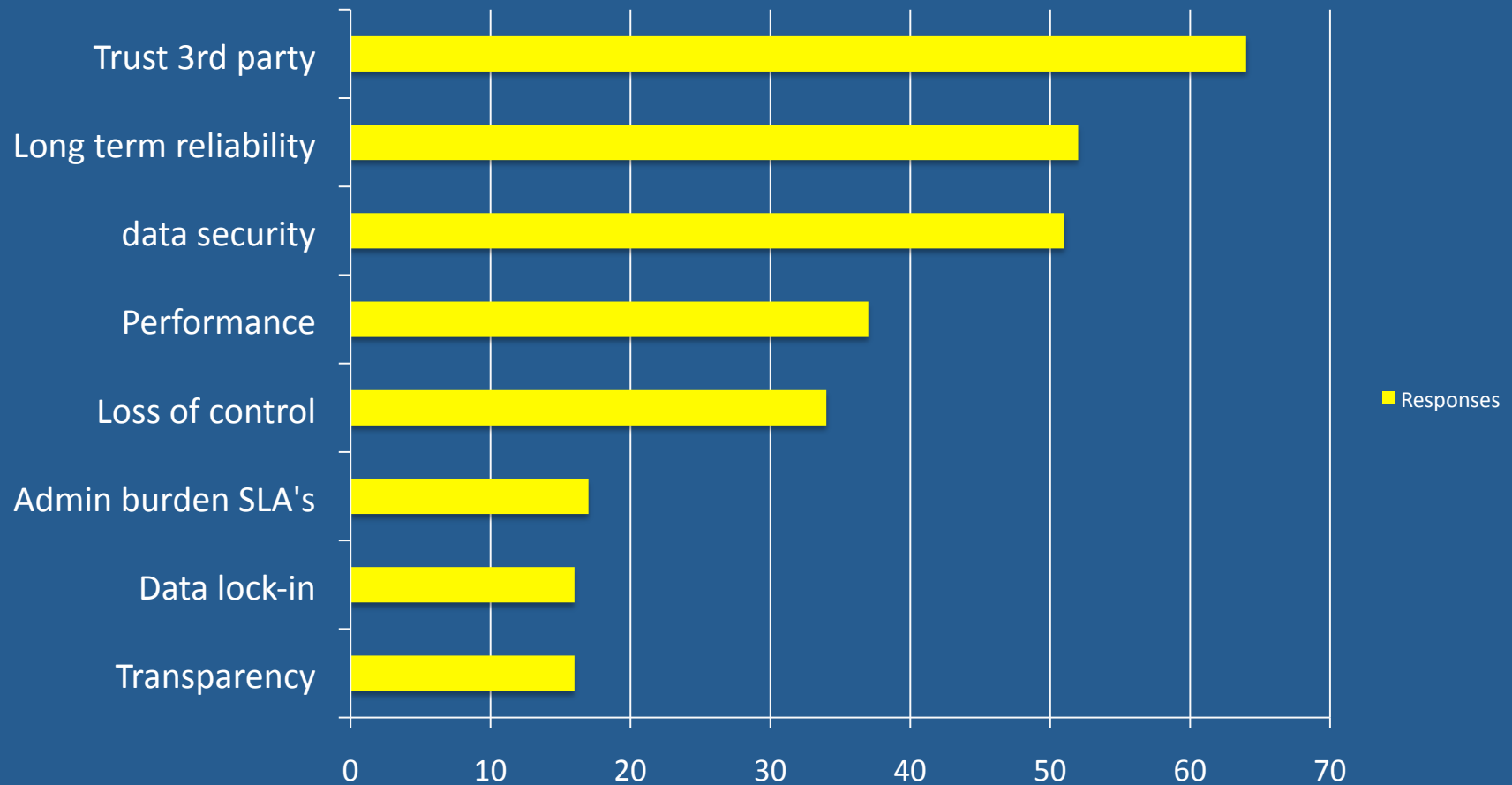
# Survey Higher Ed, 145 responses

## Key Benefits



# Survey Higher Ed, 145 responses

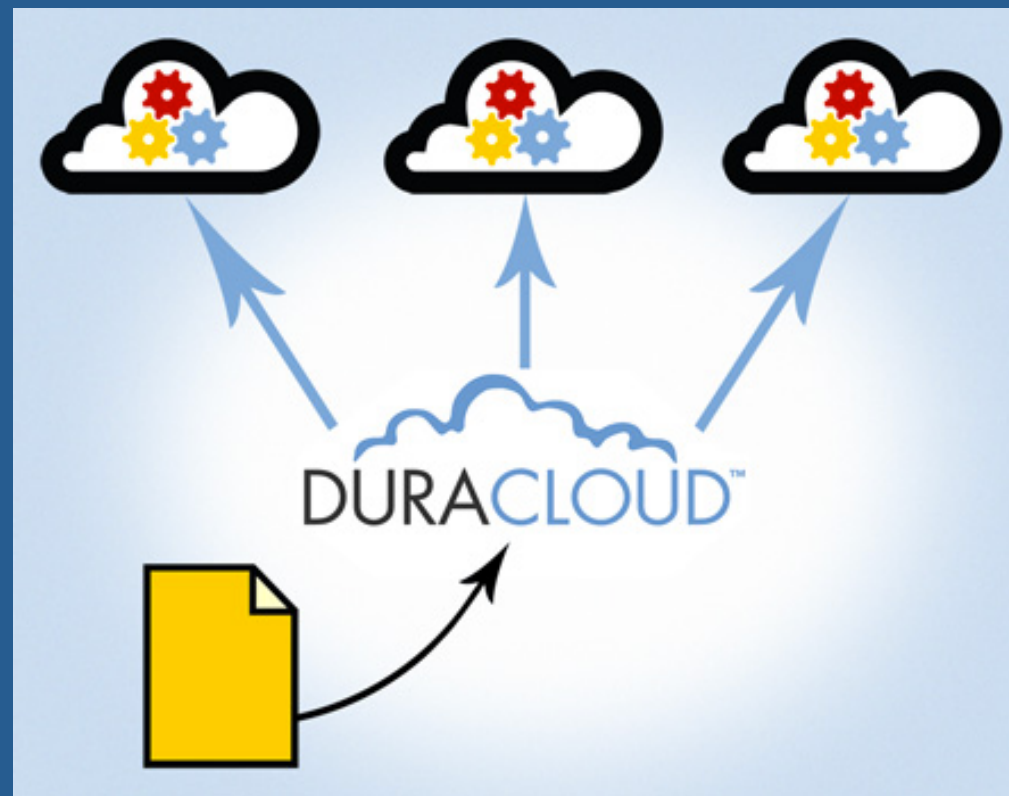
## Key Challenges



# What is DuraCloud?

Digital archiving solution based on cloud infrastructure

Across multiple cloud providers



# DuraCloud apps

## Archiving and Preservation



Online Backup(s)



File Format  
Identification



File health check



Synchronization to  
multiple clouds

## Multimedia Access



Advanced  
Image Viewing



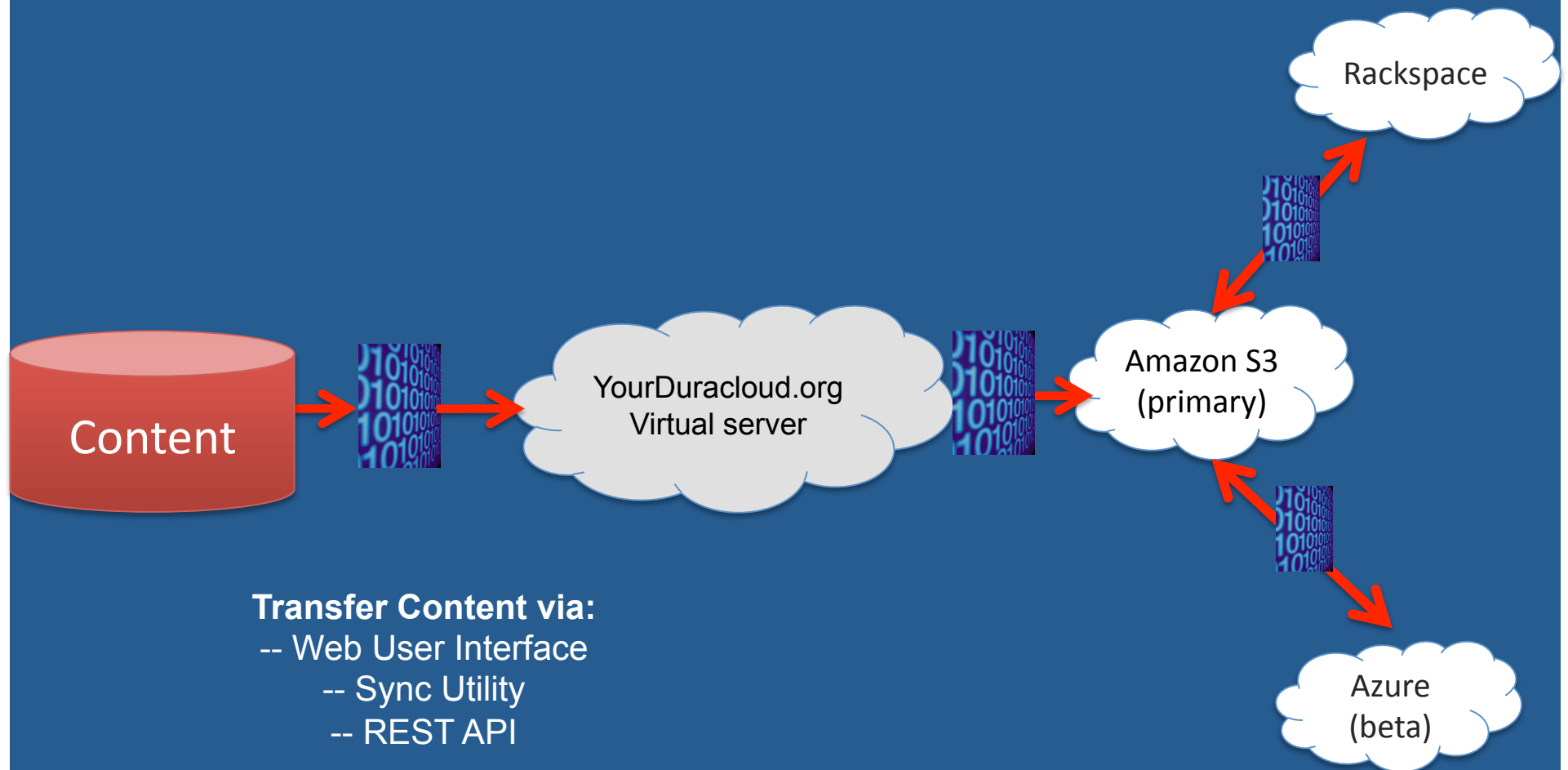
Media Streaming



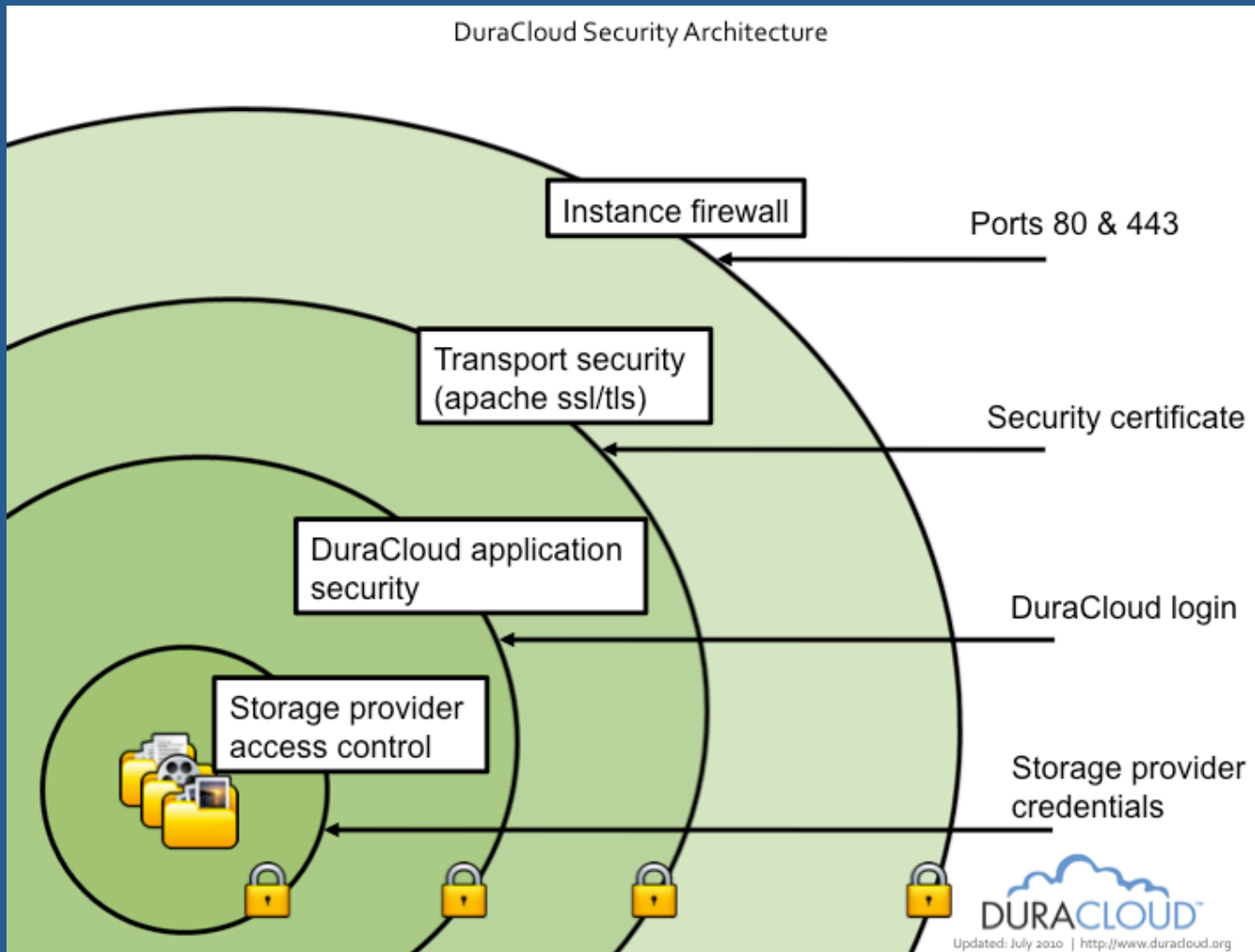
Image  
Transformation

...more on the roadmap

# How does DuraCloud work?

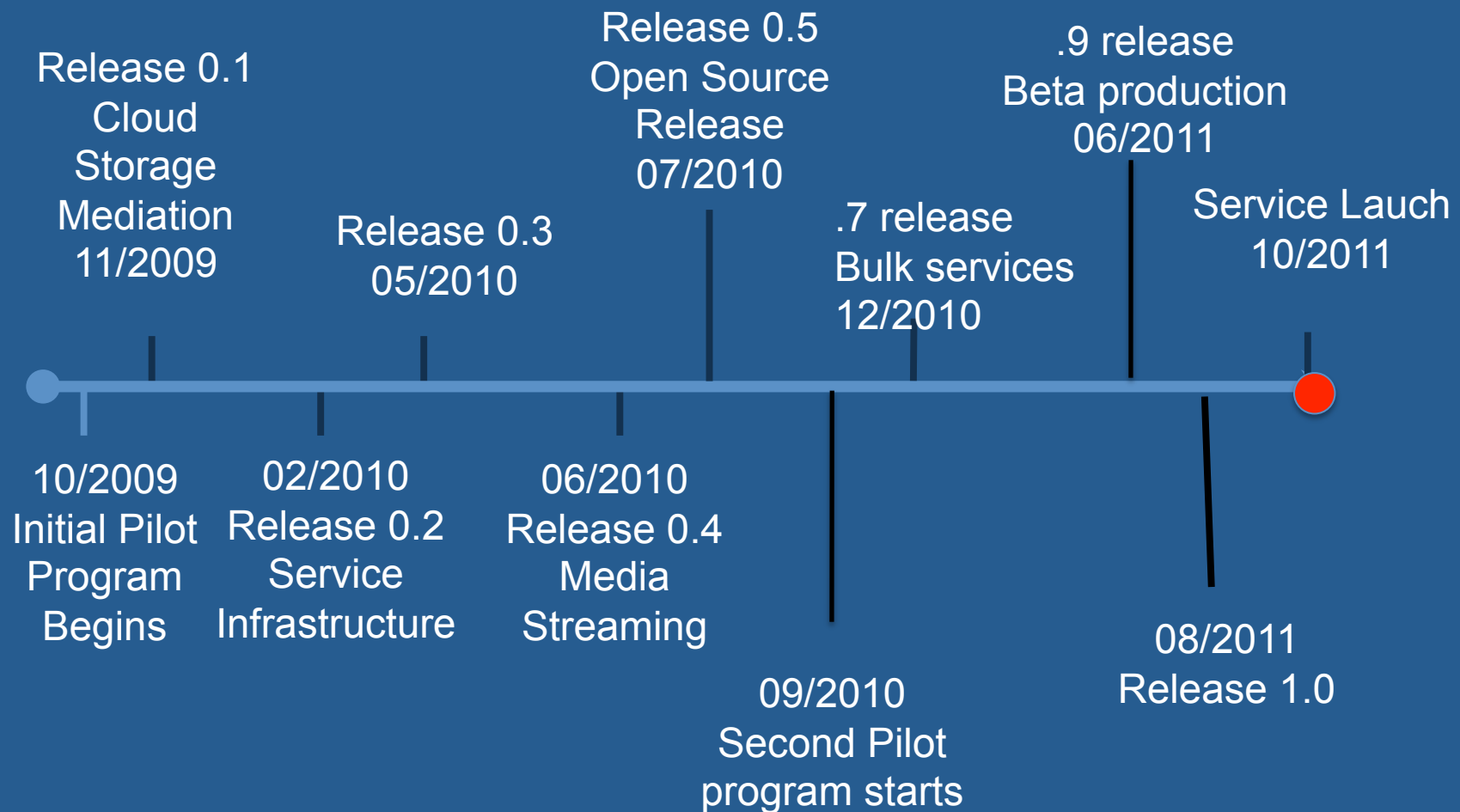


# DuraCloud security Architecture





# DuraCloud Project History



# 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 access	Fedora/Islandora
ICPSR	Preservation	Fedora
IUPUI	Preservation	DSpace, Content DM
Rhodes College	Image Access	DSpace
North Carolina State U	Preservation	DSpace
CARL	Preservation and Services	Fedora
MIT	Preservation	Dspace
Columbia	Preservation and Services	IA, Fedora

# Use cases explored

Digital archiving



Video streaming



Data management



Image serving



Collaboration

# Archiving and Preservation support

- Requirements met
  - ✓ Easy back up to primary and secondary cloud
  - ✓ Keep backups in sync
  - ✓ Check health of backups
  - ✓ Ability to view and download files
  - ✓ Retrieve and restore files
  - ✓ Web accessible

# On the roadmap

- Services/Apps
  - Improved video handling
  - File validation
  - Encryption
- Storage providers
  - SDSC Cloud
  - Azure
  - Local Eucalyptus integration
- Simplification

# Key Findings from Pilots

- Biggest challenge was transferring data from content owner
- Extremely difficult to compare costs
- Integration with local systems
- Biggest value of DuraCloud was simplicity and abstraction of individual cloud stores

# Pilot partner comments

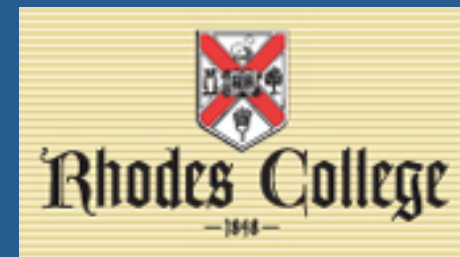
- “The ***ease-of-use*** of DuraCloud is its most impressive feature.”  
~Columbia



- “DuraCloud delivers the ***benefits*** of a diverse network of storage locations, but ***without the overhead*** of managing different vendors.”  
~ICPSR



- “One of the best things that DuraCloud does is ***bridge the complex gap*** between normal users and cloud-based services, and it does this job very well.”  
~Rhodes College



# What we learned about the cloud

- Durability- No loss of data over 2 year period, 15+ accounts, 15+ TB BUT no published SLA or policy
- Eventual consistency
- Data transfer rates over http, 1 GB realistic
- Large file handling evolving- 5 GB chunk size
- Verification of content via check sum depends on provider
- No standard API's yet across providers
- Instability of cloud providers ( Atmos,Sun)



# Reaching for the Cloud

**EDUCAUSE 2011**

Stu Baker

[stubaker@northwestern.edu](mailto:stubaker@northwestern.edu)



NORTHWESTERN  
UNIVERSITY

## On the docket

- Northwestern's computing environment
- Digital Preservation Strategies
- Duracloud pilot
- Cloud Challenges



# NORTHWESTERN UNIVERSITY LIBRARY

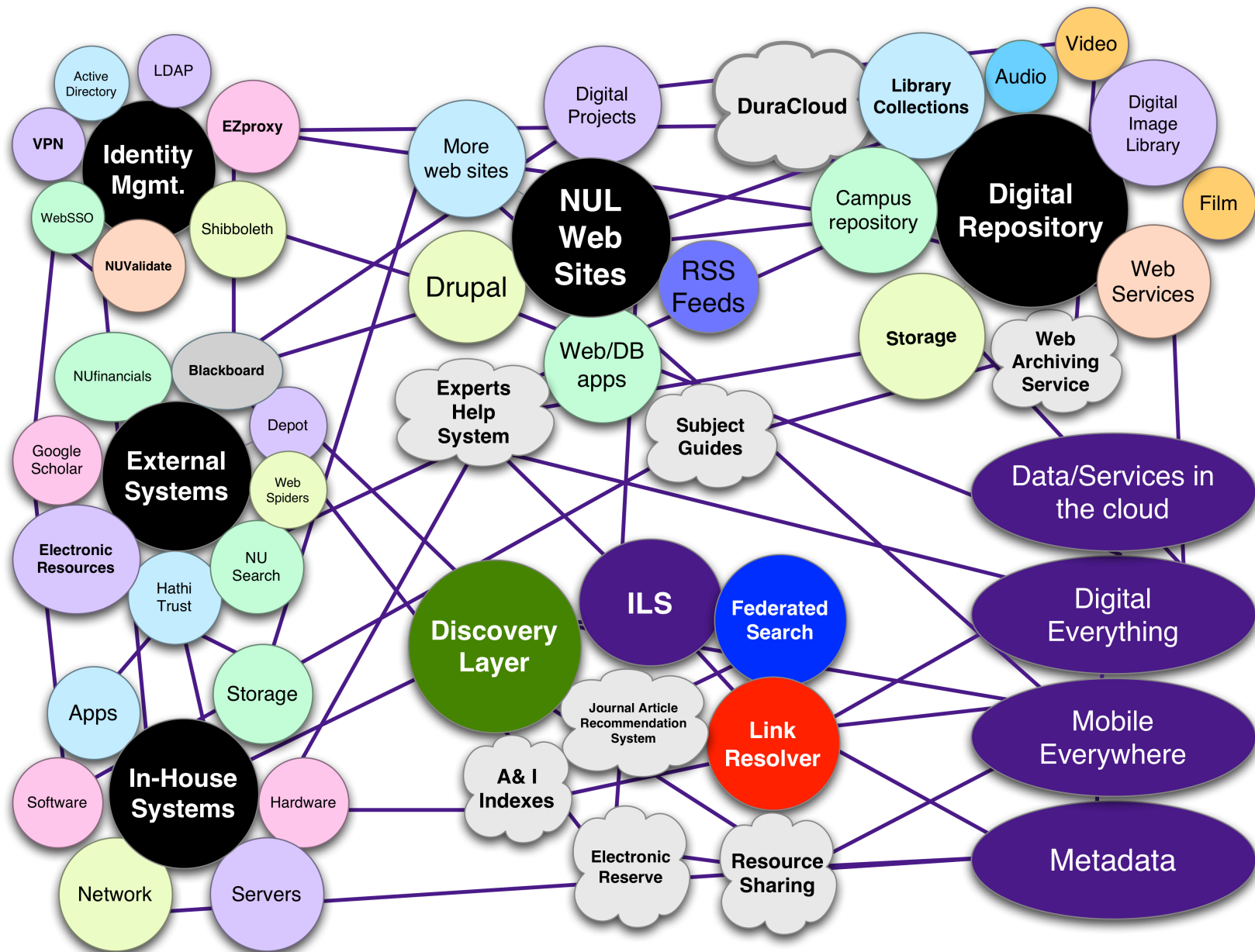


**4,930,613** Books  
**208,757** Maps  
**601,153** Images  
**66,977** Audio  
**33,882** Film and video

**98,844** Journals and serials  
**22 TB** Unique digital content  
**84** Web sites  
**750** Workstations  
**59** Servers



NORTHWESTERN  
UNIVERSITY





## Moving towards a strategy

- Preservation risk assessment, metadata extraction, format migration planning, object identification and validation
- Digital Preservation Services and Workflow
- Scalable Architecture
- Hybrid storage strategy
- Re-allocation of staff expertise and local technology



## DuraCloud pilot

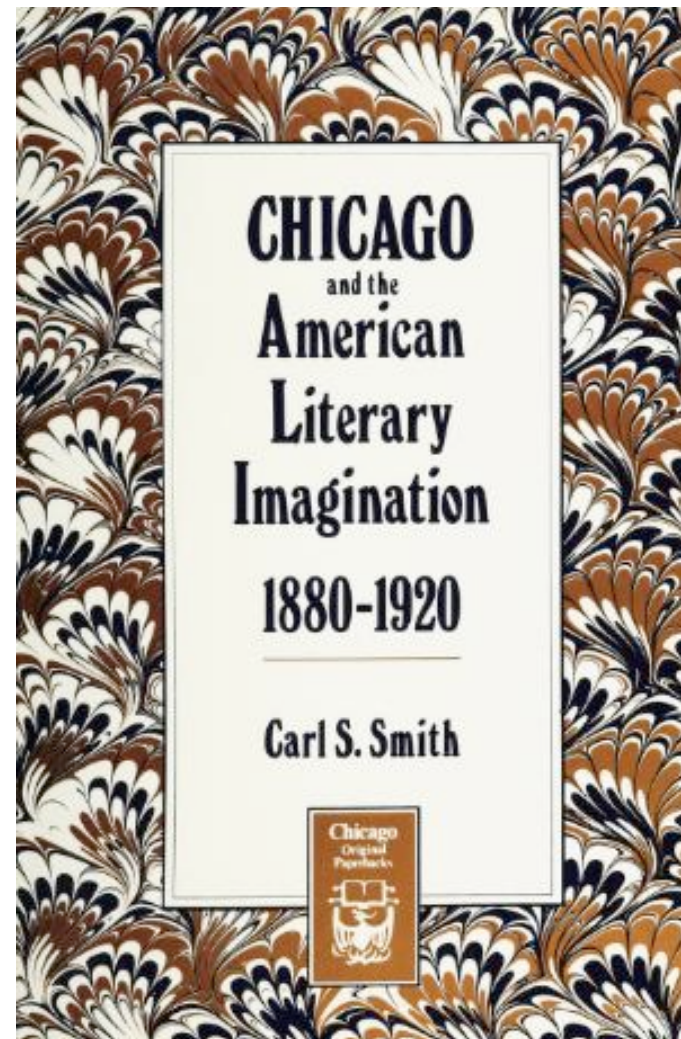
- Dark archive
- Locally digitized books
- Fedora repository

For a 300 page book:

- 300 camera jpegs
- 300 edited jpegs
- 300 jp2s
- 300 pdfs
- 300 OCR txt file
- 300 OCR XML files



NORTHWESTERN  
UNIVERSITY



## Cloud Challenges

- Service Provider Security Assessment
- Addressing PII, PHI, or other sensitive data concerns
- More sophisticated access control provisions
- Determining lifecycle costs for preservation;  
Consortial deals?
- Tiered mass storage strategy; How far do we go?



# Upcoming Webinars

- ✦ *Using DuraCloud for Archiving and Preservation*  
November 2<sup>nd</sup> at 1pm ET
- ✦ *Technical Overview of DuraCloud*  
November 16<sup>th</sup> at 1pm ET
- ✦ *DSpace and DuraCloud integration*
- ✦ *Fedora and DuraCloud integration*  
November/December days/times TBA



# Free DuraCloud Trial Accounts



The screenshot shows the DuraCloud website homepage. At the top, the DuraCloud logo is on the left, and links for "DOCUMENTATION | SUPPORT" with social media icons are on the right. Below the logo is a navigation bar with "Overview", "Solutions", "Features", and "Learn More". A search bar is positioned to the right of the navigation bar. The main banner features a colorful map background with the text "Store and do more" in large yellow letters, followed by "ON-DEMAND STORAGE AND SERVICES FOR YOUR DIGITAL CONTENT" in white. Below this, a section titled "WHAT USERS ARE SAYING" is visible. On the left side of the banner, there are three call-to-action buttons: "FIND OUT MORE", "TRY IT" (which is circled in red), and "GET STARTED". On the right side of the banner, there is a video player showing the DuraCloud logo and a play button. The video player has a progress bar at the bottom indicating "00:00 00:00".

DURACLOUD™

DOCUMENTATION | SUPPORT

Overview Solutions Features Learn More

**Store and do more**  
ON-DEMAND STORAGE AND SERVICES FOR YOUR DIGITAL CONTENT

WHAT USERS ARE SAYING

FIND OUT MORE

**TRY IT**

GET STARTED

DURACLOUD™

00:00 00:00

# Where can I find out more?

The screenshot shows the DuraCloud website homepage. At the top, the DuraCloud logo is on the left, and links for DOCUMENTATION, SUPPORT, and social media icons are on the right. Below the logo is a navigation bar with links: Overview, Solutions, Features, and Learn More. A search bar is positioned to the right of these links. The main banner features a colorful mosaic background with the text "Store and do more" in large yellow letters, followed by "ON-DEMAND STORAGE AND SERVICES FOR YOUR DIGITAL CONTENT" in white. Below this, a section titled "WHAT USERS ARE SAYING" is partially visible. On the left side of the banner, there are three circular buttons with play icons: "FIND OUT MORE", "TRY IT", and "GET STARTED". To the right of these buttons is a video player showing the DuraCloud logo and a play button. The footer contains four columns of links: "News and Events" (including a link about a Dropbox glitch), "Spotlights" (including a link about a 19th-century view of Shakespeare), "Open Source Community" (including links to Why Open Source?, Discussion Forum, Documentation, Issue Tracker, and Latest Release), and "Customer Help and Support" (including links to Contact Us, Pricing, Newsletter Sign-Up, and Documentation). At the bottom, it states "The DuraCloud open technology and hosted service is brought to you by DURASPACE" and "Powered by Fedora, Islandora and Discovery Garden | Licensed under Creative Commons Attribution-Share Alike 3.0 Unported License | Site Map".

- Web site:  
[www.duracloud.org](http://www.duracloud.org)
- Email:  
[csmith@duraspace.org](mailto:csmith@duraspace.org)

# The open source platform

- Available for organizations to run locally
- Requires cloud ( commercial or private) infrastructure
- Being tested by several organizations now
- Looking for organizations that want to enable their “software as a service” in DuraCloud
- For more details visit the wiki at:  
<https://wiki.duraspace.org/display/duracloud/>