



DURACLOUD™

Archiving and Preservation

Michele Kimpton
CEO, DuraSpace

Bryan Beecher
Director, ICPSR

DuraSpace Webinar
November 2, 2011



DuraSpace Mission

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

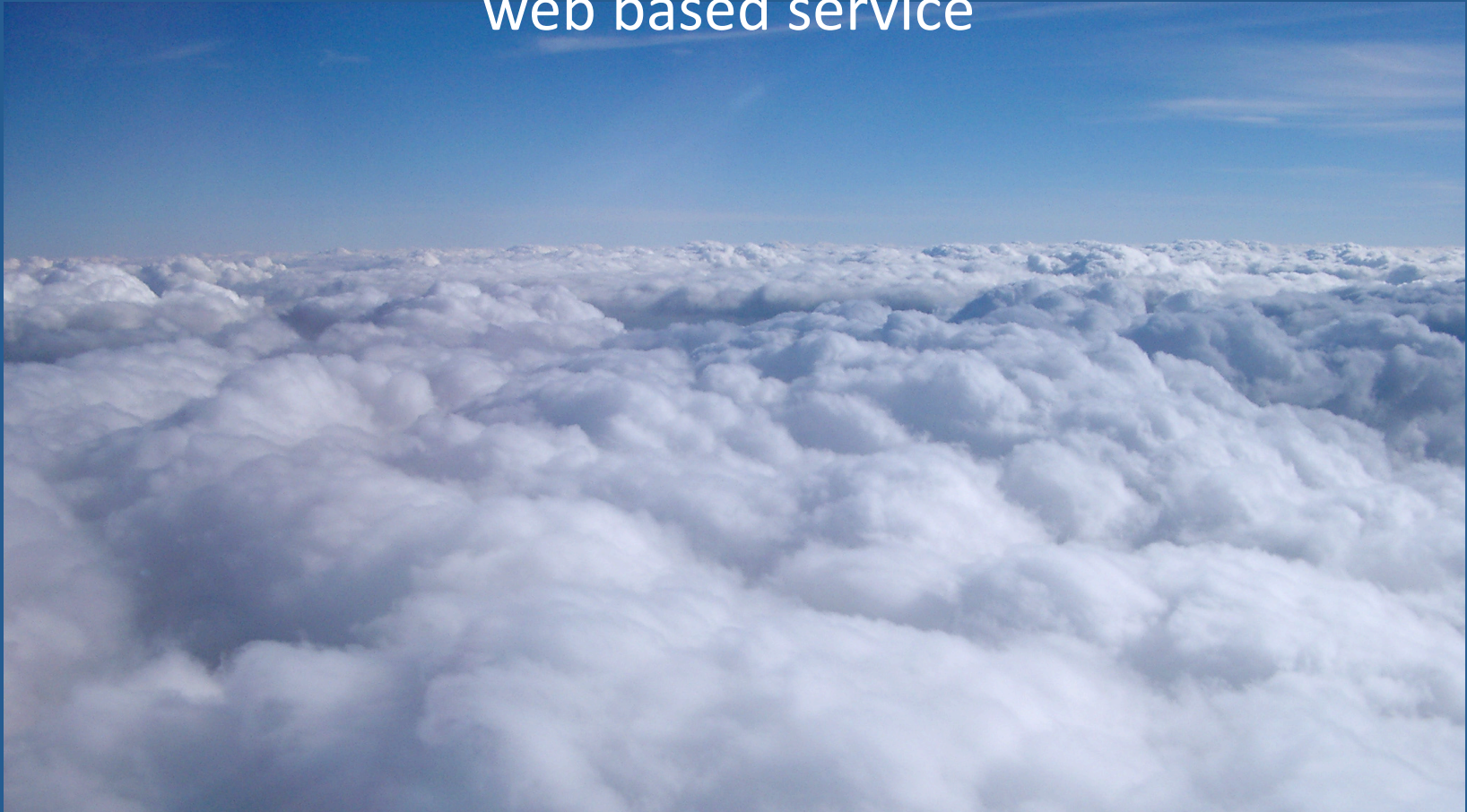


Preservation challenges

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

Why cloud?

Massively scalable compute and storage offered as a web based service



Higher Ed survey, 211 responses

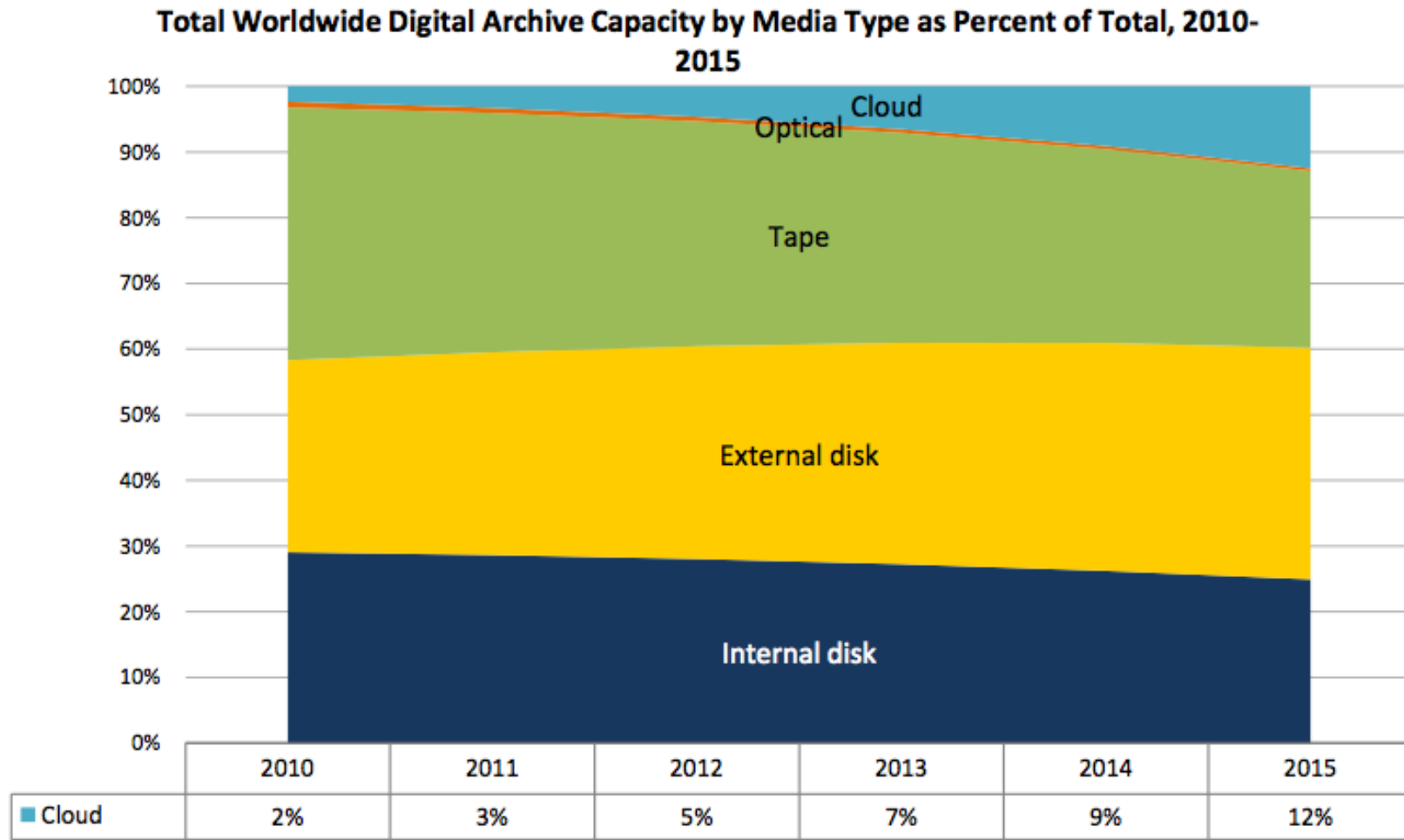
8. How likely are you to consider cloud services for the following services?

Answer

	Very likely	Likely	Unlikely	Very unlikely
Online primary storage of digital content	18.0% (34)	24.9% (47)	28.6% (54)	20.1% (38)
Online back-up of digital content: a remote, secondary store of digital content	29.0% (54)	46.8% (87)	14.5% (27)	4.3% (8)
Preservation support: content replication, auditing, repair	19.9% (37)	43.5% (81)	16.1% (30)	5.9% (11)
Shared infrastructure for managing digital content within the institution	16.8% (31)	38.4% (71)	23.8% (44)	9.2% (17)
Shared infrastructure for managing digital content across institutions	16.8% (31)	39.5% (73)	18.4% (34)	10.3% (19)
Ability to provision compute services on stored digital content	12.6% (23)	30.8% (56)	25.3% (46)	5.5% (10)

Digital archiving by media type

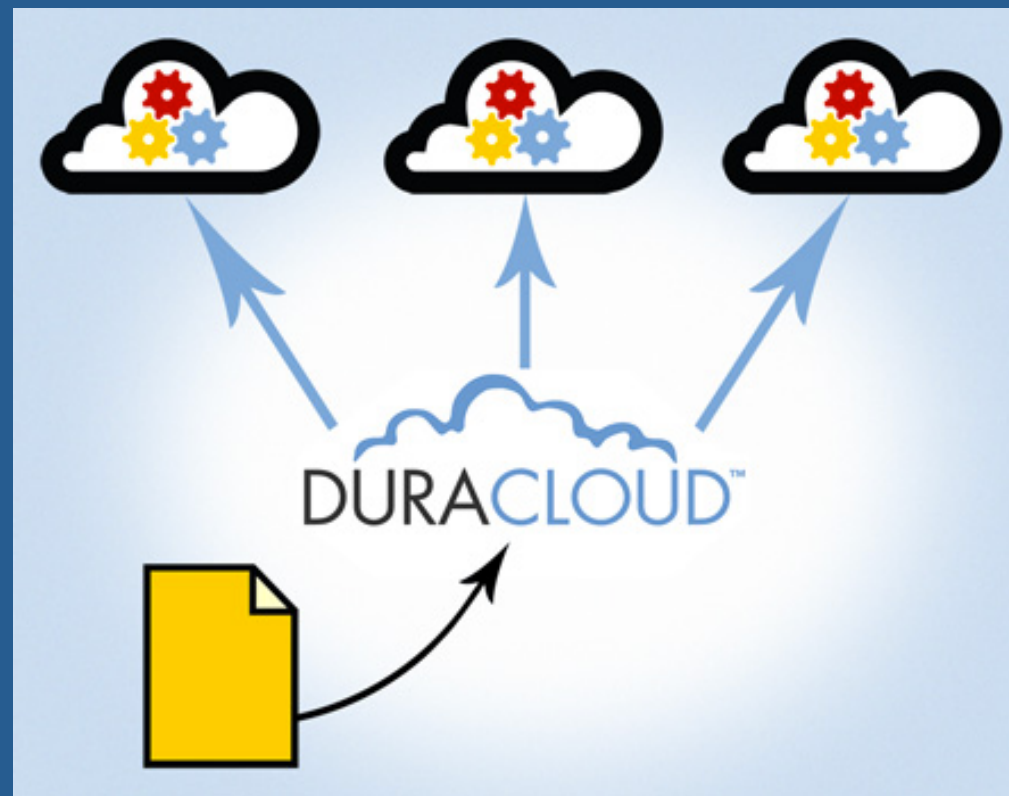
Figure 2. Total Worldwide Digital Archive Capacity, by Media Type as Percent of Total, 2010-2015



What is DuraCloud?

Platform and service based on cloud infrastructure

Across multiple cloud providers



DuraCloud apps

Archiving and Preservation focused-



Online Backup(s)



File Format
Identification



File health check



Synchronization of
content to multiple
clouds

...more on the roadmap

Archiving and Preservation support

- Duracloud provides
 - ✓ Easy back up to multiple cloud providers
 - ✓ Keep backups in sync
 - ✓ Check health of backups
 - ✓ Ability to view and download files
 - ✓ Retrieve and restore files
 - ✓ Web accessible

Using DuraCloud for Archiving & Preservation

Bryan Beecher
Director, Computer &
Network Services
ICPSR



ICPSR
A PARTNER IN
SOCIAL SCIENCE
RESEARCH

50
years

About ICPSR

- Inter-university Consortium for Political and Social Research
- Located at the University of Michigan
- World's largest archive of social science research data
- In operation for 50 years
- About \$15m in revenues

Archival holdings

- Lots of little files
 - text/plain
 - application/pdf
 - text/xml
 - other stuff
- 2m files; 6TB of storage

Strategy

- Bit-level for original (SPSS + Word)
- Normalize into more durable formats (plain text data + XML metadata + PDF/A documentation)
- Transform for better delivery
- Retain transform and derivatives
- Lots of copies

Data archiving, 1 BC



ICPSR
A PARTNER IN
SOCIAL SCIENCE
RESEARCH

50
years

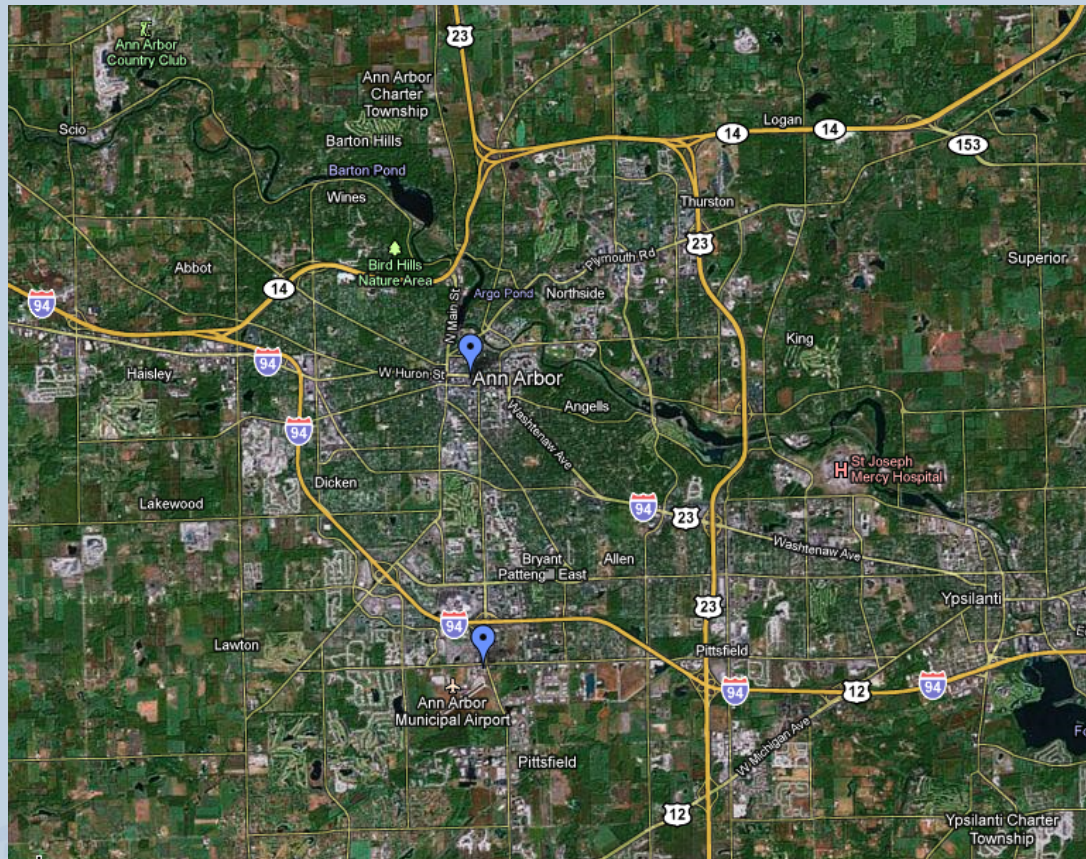
Geographic Diversity, 1 BC



Geographic Diversity, 1 BC



Geographic Diversity, 1 BC

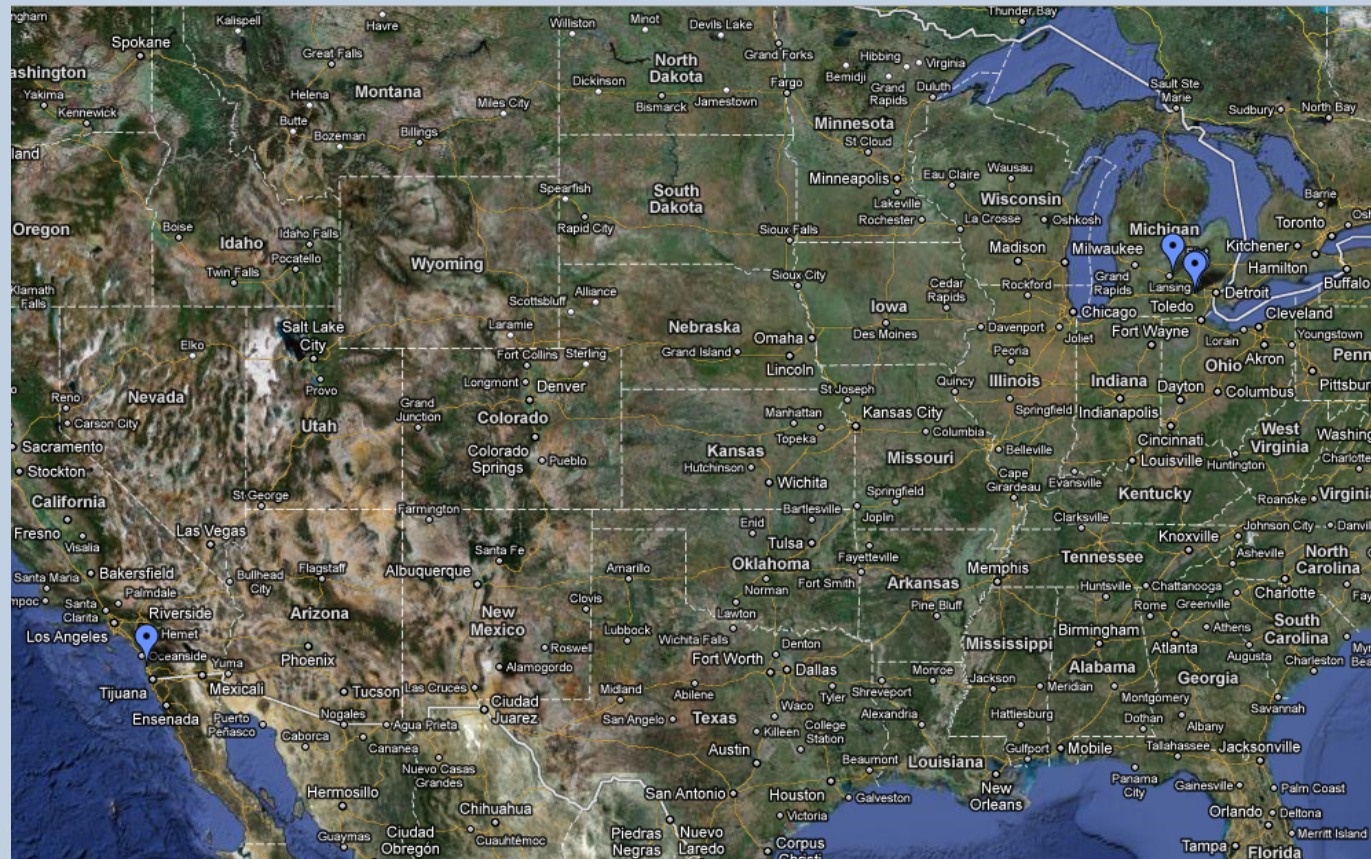


Maybe disk instead of tape?



- Synchronize content to other locations
- Fixity checking lets us know when we need to “fix” something

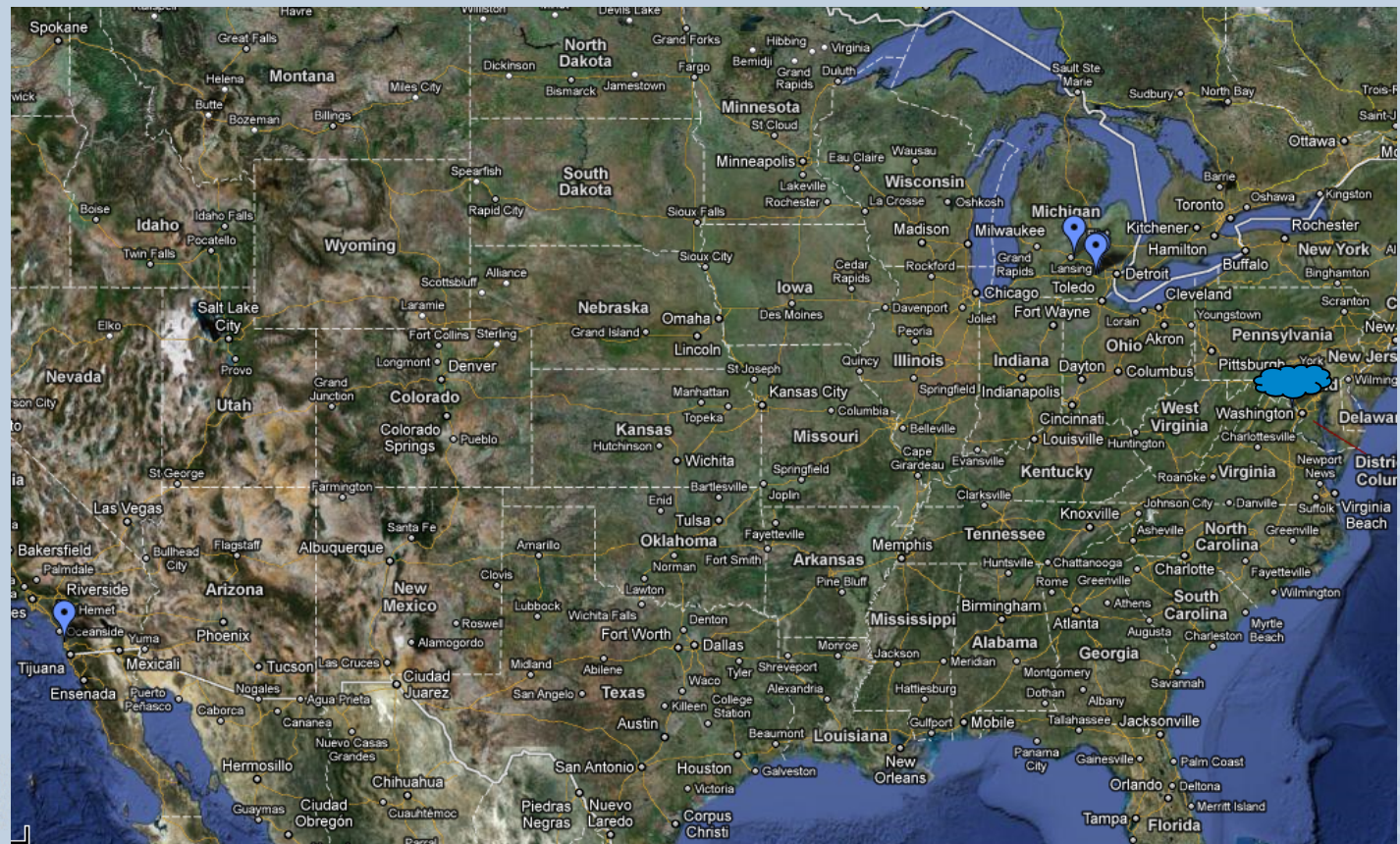
Get by with a little help from our friends



And they are friends

- Based on relationships
- No SLA
- No scale up/down
- Idiosyncratic interface
- *Contracts? We don't need no stinkin' contracts!*

A copy in the cloud



Are you crazy?

- FISMA Low
 - Not encrypted
 - Machine room open access
 - Firewalled
 - Professional IT staff + others
- FISMA Medium
 - Encrypted
 - Machine room controlled access
 - Firewalled
 - Professional IT staff

Honeymoon period

- Automated monthly billing for usage (storage, computer, network I/O)
 - Small EC2 instance + 6 x 1TB EBS volumes bound together as a RAID
- Easy to scale up and down
- Easy to synchronize

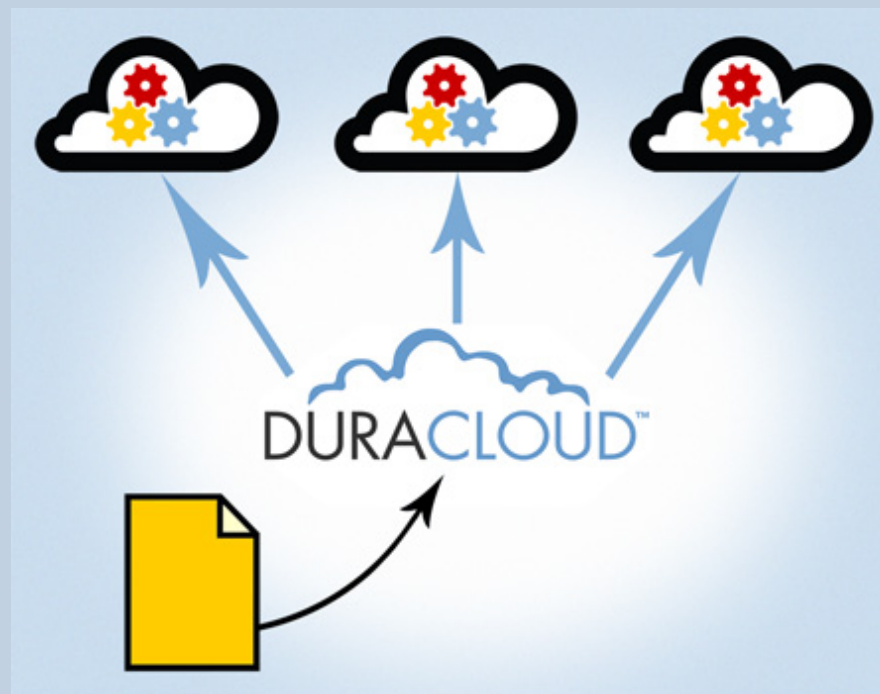
And best of all...



So what's not to like?

- Cloud diversity
 - Location
 - Technology platform
 - Operational processes
 - Business viability
- Vendor lock-in

Who can save us?



What we like

- Single interface to “the cloud”
- Single billing contact
 - Single relationship
- Value-added services
 - Fixity checking

What we would change

- Filesystem semantics would work better for us
 - *rsync* v. *synctool*
 - files v. objects
- Support for big files/objects
- Tools suitable for automated batch use (i.e., out of *cron*)

Takeaways

- Cloud is a viable option for additional archival copies
- Physical infrastructure may be at least as good as your own
- Encrypt the sensitive stuff
- Not the low-cost solution; but may be the low-hassle solution

More info

- Bryan Beecher
 - bryan@umich.edu
 - <http://techaticpsr.blogspot.com/>

Upcoming DuraCloud Webinars

✦ *Technical Overview of DuraCloud*

November 16 at 1pm ET

✦ *DSpace and DuraCloud*

November 30 at 1pm ET


✦ *Fedora and DuraCloud*

January 11 at 1pm Et

Try DuraCloud Free for One Month: Trial or Subscription

The screenshot shows the DuraCloud website homepage. At the top, the DuraCloud logo is on the left, and navigation links for 'MY ACCOUNT', 'DOCUMENTATION', 'SUPPORT', and social media icons are on the right. Below the logo is a menu with 'Overview', 'Solutions', 'Features', and 'Learn More'. A search bar is positioned to the right of the menu. 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, there are three buttons: 'FIND OUT MORE', 'TRY IT' (circled in red), and 'SUBSCRIBE'. At the bottom left, a yellow circular badge with a red border says 'get one month FREE'. On the right, there is a video player showing the DuraCloud logo with a play button overlay. The video player controls at the bottom show a progress bar and a timestamp of '00:00 00:00'.

DURACLOUD™

MY ACCOUNT | 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

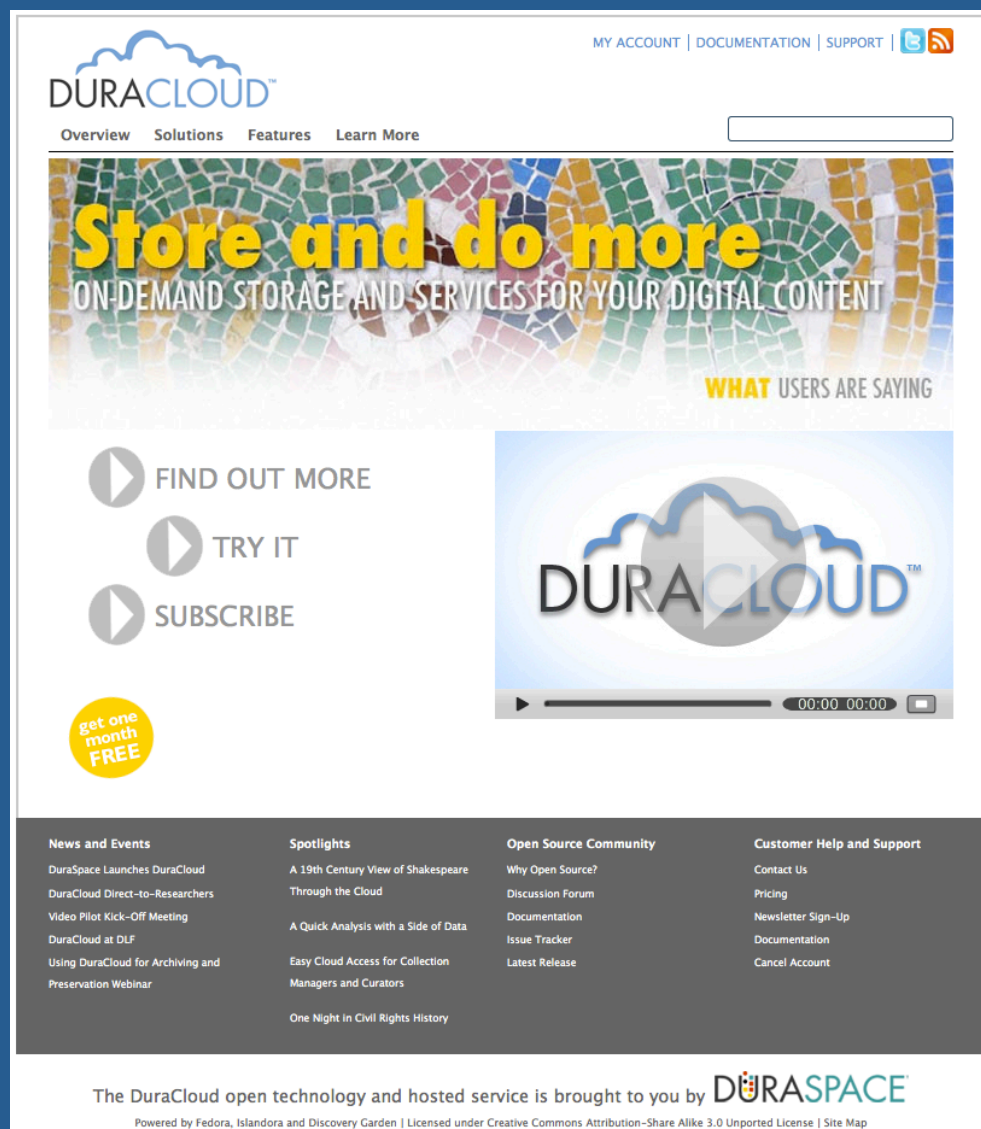
SUBSCRIBE

get one month FREE



DURACLOUD™

00:00 00:00

Where can I find out more?



The screenshot shows the DuraCloud website homepage. At the top, there's a navigation bar with links for 'MY ACCOUNT', 'DOCUMENTATION', 'SUPPORT', and social media icons. Below this is the DuraCloud logo and a search bar. 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'. A video player is embedded in the center, showing the DuraCloud logo and a play button. To the left of the video are three buttons: 'FIND OUT MORE', 'TRY IT', and 'SUBSCRIBE'. A yellow circular badge on the left says 'get one month FREE'. The footer contains four columns of links: 'News and Events', 'Spotlights', 'Open Source Community', and 'Customer Help and Support'. At the bottom, it states 'The DuraCloud open technology and hosted service is brought to you by DURASPACE' and provides licensing information.

MY ACCOUNT | DOCUMENTATION | SUPPORT |  

DURACLOUD™

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

SUBSCRIBE

get one month FREE

DURACLOUD™

00:00 00:00

News and Events
DuraSpace Launches DuraCloud
DuraCloud Direct-to-Researchers
Video Pilot Kick-Off Meeting
DuraCloud at DLF
Using DuraCloud for Archiving and Preservation Webinar

Spotlights
A 19th Century View of Shakespeare Through the Cloud
A Quick Analysis with a Side of Data
Easy Cloud Access for Collection Managers and Curators
One Night in Civil Rights History

Open Source Community
Why Open Source?
Discussion Forum
Documentation
Issue Tracker
Latest Release

Customer Help and Support
Contact Us
Pricing
Newsletter Sign-Up
Documentation
Cancel Account

The DuraCloud open technology and hosted service is brought to you by **DURASPACE™**
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
- Email:
csmith@duraspace.org