DURACLOUD™
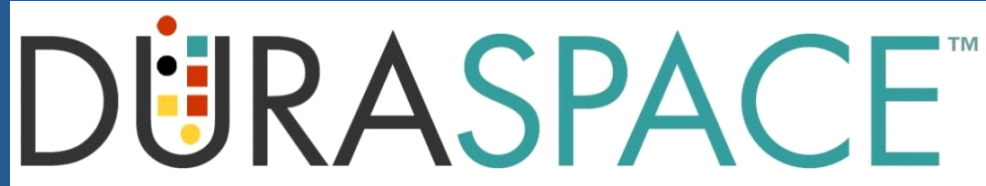
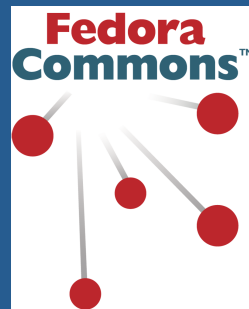## Archiving and preserving in the cloud, challenges and benefits

Michele Kimpton, DuraSpace
CEO DuraSpace

"DuraSpace is an independent 501(c)(3) not-for-profit organization providing leadership and innovation for open technologies that promote durable, persistent access to digital data. We collaborate with academic, scientific, cultural, and technology communities by supporting projects and creating services to help ensure that current and future generations have access to our collective digital heritage."

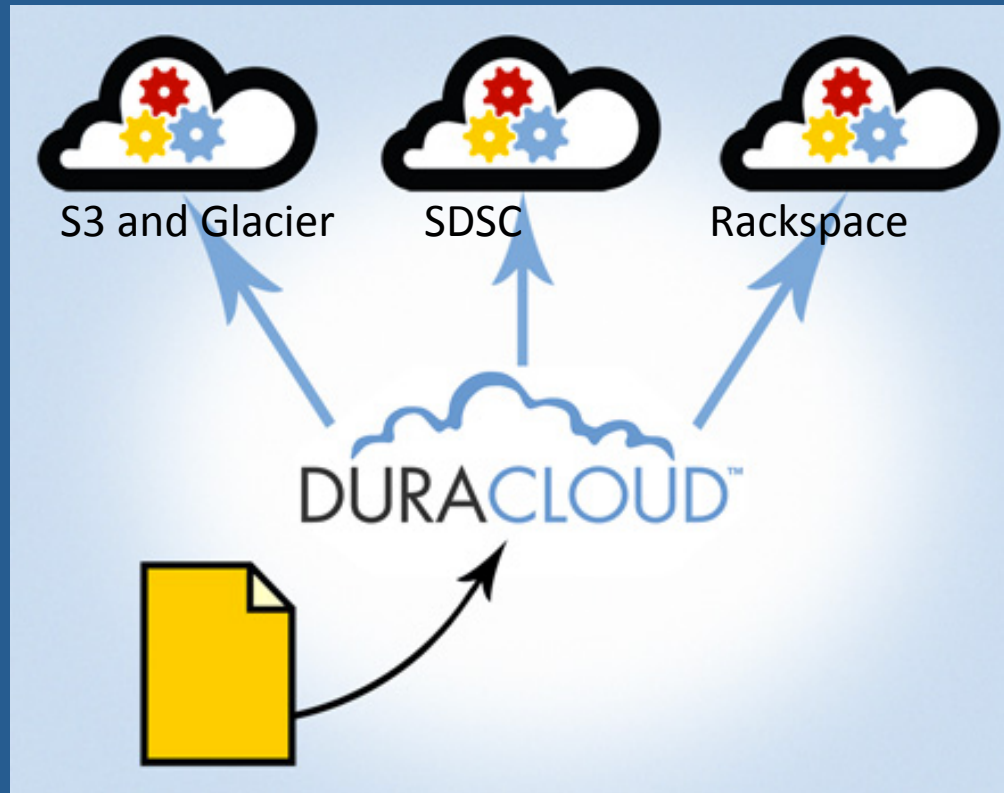# DuraSpace projects and services

Projects

Services

What is DuraCloud?

Archiving and preservation services in the cloud

Ability to choose one or multiple cloud storage providers

Why we chose the cloud?

- **STABLE**-Open API's(application program interface) easy to use and stable

- **FLEXIBLE**-Flexible, scale up and down as you need

- **EASE OF USE**-Takes a few hours(at most) to spin up an account and start ingesting data- IT support not needed

- **ACCESSIBILITY**- Online 24/7 with web access

- **COST**-S3 has reduced its price 65% in the last 3 years (12 cents to 3 cents per gb/mo)

Challenges

- SLA-No Durability guarantees, everything based on performance/access

- Cost model-Institutions prefer fixed TB cost- treat like a capital expenditure

- Transparency-No transparency on data handling policies

- Data Integrity-No auditable health checking

- Data transfer-difficult to move large data sets over the wire consistently (more the 2GB files sizes is tough)

- Privacy- must encrypt for assurance, not HIPPA compliant

    Majority of issues are business issues not technical issues

POWRR
Preserving (Digital) Objects With Restricted Resources

Digital POWRR Tool Evaluation Grid	Ingest					Processing						Access		Storage					Maintenance			Other		
	Copy	Fixity Check	Virus Scan	File Dedupe	Auto Unique ID	Auto Metadata Creation	Auto Metadata Harvest	Manual Metadata	Rights Management	Package Metadata	Auto SIP Creation	Public Interface	Auto DIP Creation	Auto AIP Creation	Reliable, Long-Term Bit Preservation	Redundancy	Geographically Dispersed Data Storage Model	Exit Strategy	Migration	Monitoring	Auto Recovery	Open Source	Clear Documentation	Cost
Dioscuri																						x	x	Free
DROID (Digital Record Object Identification)																						x	x	Free
Dropbox	x			x								x				x	x	x					x	Varies
DSpace Direct	x	x	x		x	x	x	x	x	x	x	x	x	x	x		x	x		x	x	x	x	Varies
DSPS (Digital Preservation Software Platform)	x	x	x	x	x	x				x				x	x	x		x	x	x		x	x	Free
Duke Data Accessioner	x	x			x	x	x	x		x	x											x		Free
**DuraCloud	x	x		x	x	x	x	x			x		x	x	x	x		x		x	x	x	x	Varies
EMET (Embedded Metadata Extraction Tool)							x															x	x	Free
EXIF to DC XML Normalizer						x																x	x	Free
Fedora3/Islandora7	x	x		x	x	x	x	x	x	x	x	x		x	x	x		x		x		x	x	Free

In production for three years

- Some stats-
  - 60+ TB of data
  - 100% Durability, no file failures
  - 15+ million items stored
  - 60+ identified file types
  - No data breached

Thank-you!

- Web:
  - http://www.duracloud.org
- Documentation:
  - http://wiki.duraspace/display/duracloud
- Webinars and Training Videos:
  - http://www.youtube.com/user/duracloudvideos
- Email:
  - csmith@duraspace.org or mkimpton@duraspace.org