archivematica®

+

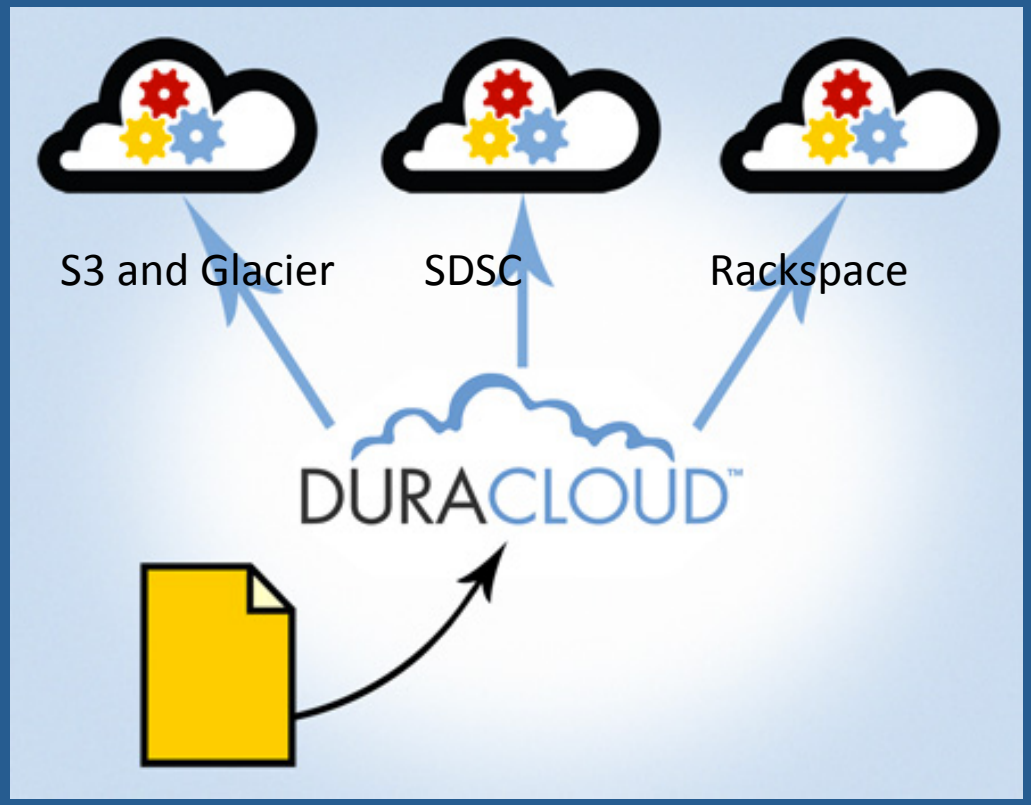
DURACLOUD™

Can these Open Source products be integrated and meet all the needs of POWRR?

Michele Kimpton, DuraSpace
CEO DuraSpace

What is DuraCloud?

Open Source Archiving and preservation platform runs in the cloud



DURASPACE™

DURACLOUD™

Features

- Can replicate and manage content to multiple cloud storage locations through single dashboard
- Automated health checking and reporting(independent of the storage provider)
- All content accessible 24/7 through web interface
- Easy to use ingest tools
- No proprietary packaging, no lock in
- Built on Open source software for full transparency

DURASPACE™

DURACLOUD™

What DuraCloud does not do

- Not a repository (like Fedora, DSpace, Islandora)
- Does not create Archival Information Packages(this needs to be done before ingest)
- No public facing portal to showcase content
- Does not produce access derivatives
- No emulation, or migration

These features typically provided by other applications integrated with DuraCloud

DURASPACE™

DURACLOUD™

What is Archivematica?

- Archivematica is a free and open-source digital preservation system designed to maintain standards-based, long-term access to digital objects
- Archivematica allows users to process digital objects from ingest to access in conformance with the ISO-OAIS functional model
- Archivematica is designed to output high-quality, standards-compliant Archival Information Packages (AIPs)
  - Bagit, METS, PREMIS

DURASPACE™ DURACLOUD™

Archivematica micro-services

- Approve transfer
- Assign file UUIDs and checksums
- Verify transfer checksums
- Generate METS.xml
- Quarantine
- Scan for viruses
- Clean up names
- Identify file format
- Validate format
- Characterize and extract metadata
- Index transfer
- Create SIP
- Verify SIP compliance
- Add metadata
- Normalize
- Process submission documentation
- Prepare DIP
- Upload DIP
- Prepare AIP
  - Generate AIP METS file
  - Index METS file
  - Bag AIP
  - Compress AIP
  - Create AIP pointer file
- Store AIP

DURASPACE™

DURACLOUD™

POWRR

Preserving (Digital) Objects With Restricted Resources

About POWRR | POWRR Blog | POWRR Presence | DP 101 | I Survived a POWRR Workshop | Contact Us | Tool Grid

Digital POWRR Tool Evaluation Grid	Ingest					Processing						Access		Storage					Maintenance			Other		
	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
ACE (Audit Control Environment)		x																		x		x		Free
AFF Open Source Computer Forensics Software							x															x		Free
Amazon S3	x	x			x	x									x	x	x	x		x	x		x	Varies
Archive-It	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x		x		x	x	Varies
**Archivematica	x	x	x		x	x	x	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	x	Free

DURASPACE DURACLOUD™

Both of these projects are Open Source

- Community driven
- Sustainability
- Transparency
- Open formats
- No vendor lock in
- Community Governance

DURASPACE™

DURACLOUD™

Pilot partners

- Illinois Wesleyan University
- State Archives North Carolina
- University of Texas
- Huntington Library
- Phillips Academy
- U of Washington Libraries
- Pepperdine University
- Berea College
- Kansas State University Libraries

DURASPACE™

DURACLOUD™

Use cases

- Each institution has selected content
- Identified accompanying materials (metadata, submission documentation, access files, service files)
- Identified materials location
- Order to files within set
- Workflow into Archivematica determined

DURASPACE™

DURACLOUD™

Observations

- Truly diverse sets of content
- Many common use cases: i.e., “Institutional photos and communications”
- Lots of descriptive metadata but from variety of sources(systems, spreadsheets)
- On multiple media throughout institution
- Few have submission documentation
- Little IT support

DURASPACE™

DURACLOUD™

Types of content

Audio, video, books, periodicals, dissertations, government publications, researcher site, websites, facebook pages, flickrr collection, content dm objects, photographs, lectures

DURASPACE™

DURACLOUD™

Steps

- ✓ Define use cases, select content
- ✓ Upload content into DuraCloud
- Process through Archivematica application
- Create AIP packages
- Deposit content into DuraCloud
- Retrieve AIP package for download and review
- Test application in accordance with POWRR grid

DURASPACE™

DURACLOUD™

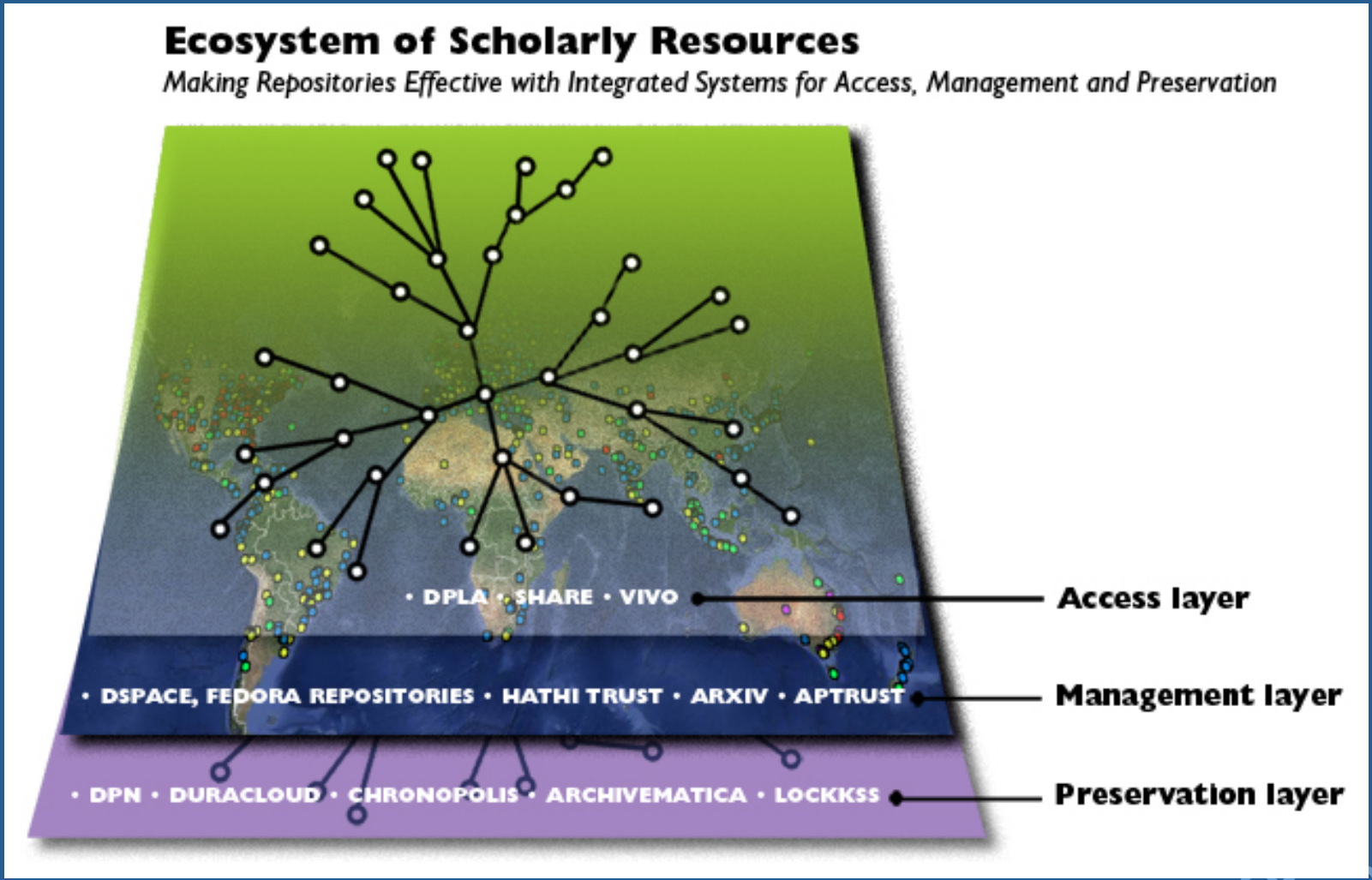
Output

- Publication of use cases tested and workflow established for variety of content types
- Does the Archivematica/DuraCloud software package meet all the elements of the POWWR grid?
- Is the software flexible enough to meet all use cases?
- What are the limitations if the software is hosted?
- Which use cases are “out of scope”?

DURASPACE™

DURACLOUD™

Connecting our projects within the ecosystem



DURASPACE™

DURACLOUD™

Thank-you!

- Web:
  - www.duracloud.org
  - www.Archivematica.org
- POWRR: http://digitalpowrr.niu.edu/tool-grid/
- Documentation
  - http://wiki.duraspace/display/duracloud
  - https://www.archivematica.org/wiki/User_Manual
- Email: mkimpton@duraspace.org
- For more info sign up: www.duracloud.org/archivematica

DURASPACE™

DURACLOUD™