



ACADEMIC PRESERVATION TRUST

Update at CNI, April 14, 2015

Chip German, APTrust

Andrew Diamond, APTrust

Nathan Tallman, University of Cincinnati

Linda Newman, University of Cincinnati

Jamie Little, University of Miami

APTrust Web Site: aptrust.org



ACADEMIC PRESERVATION TRUST

Sustaining Members:

Columbia University
Indiana University
Johns Hopkins University
North Carolina State University
Penn State
Syracuse University
University of Chicago
University of Cincinnati

University of Connecticut
University of Maryland
University of Miami
University of Michigan
University of North Carolina
University of Notre Dame
University of Virginia
Virginia Tech



ACADEMIC PRESERVATION TRUST

- Quick description of APTrust: The Academic Preservation Trust (APTrust) is a consortium of higher education institutions committed to providing both a preservation repository for digital content and collaboratively developed services related to that content. The APTrust repository accepts digital materials in all formats from member institutions, and provides redundant storage in the cloud. It is managed and operated by the University of Virginia and is both a deposit location and a replicating node for the Digital Preservation Network (DPN).
- 39 of the registrants for CNI in Seattle are from APTrust sustaining-member institutions

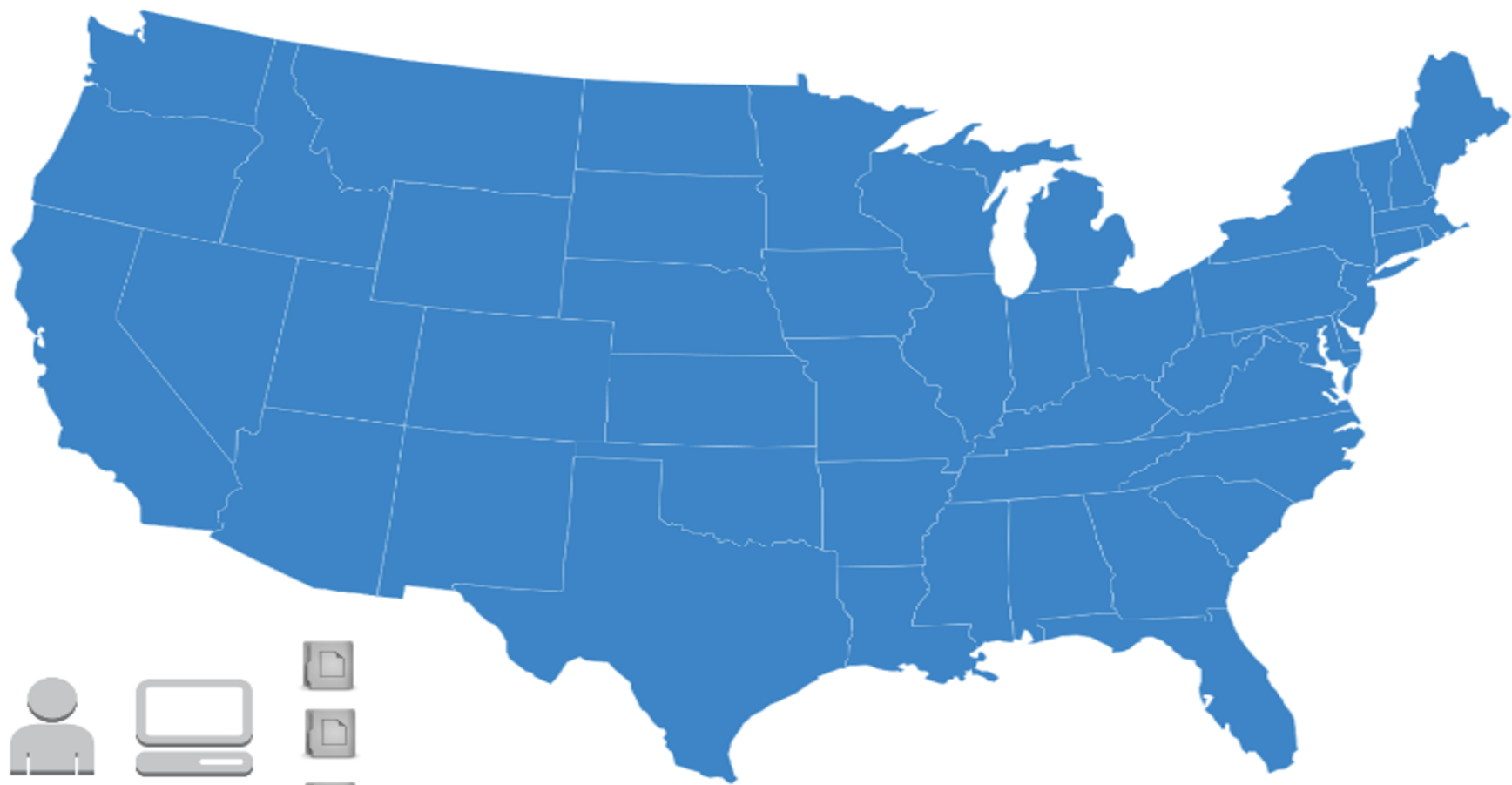


ACADEMIC
PRESERVATION TRUST

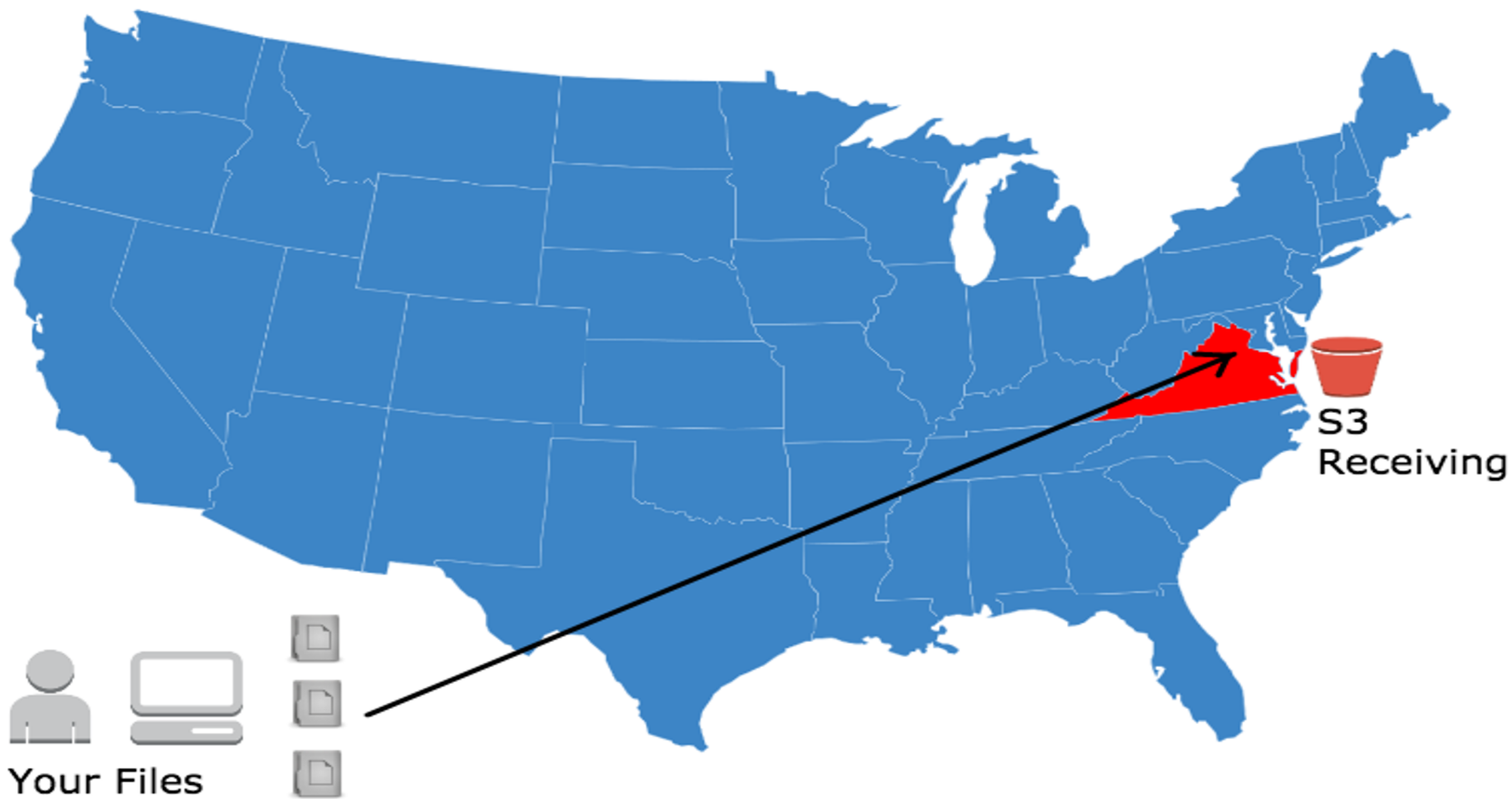
APTrust

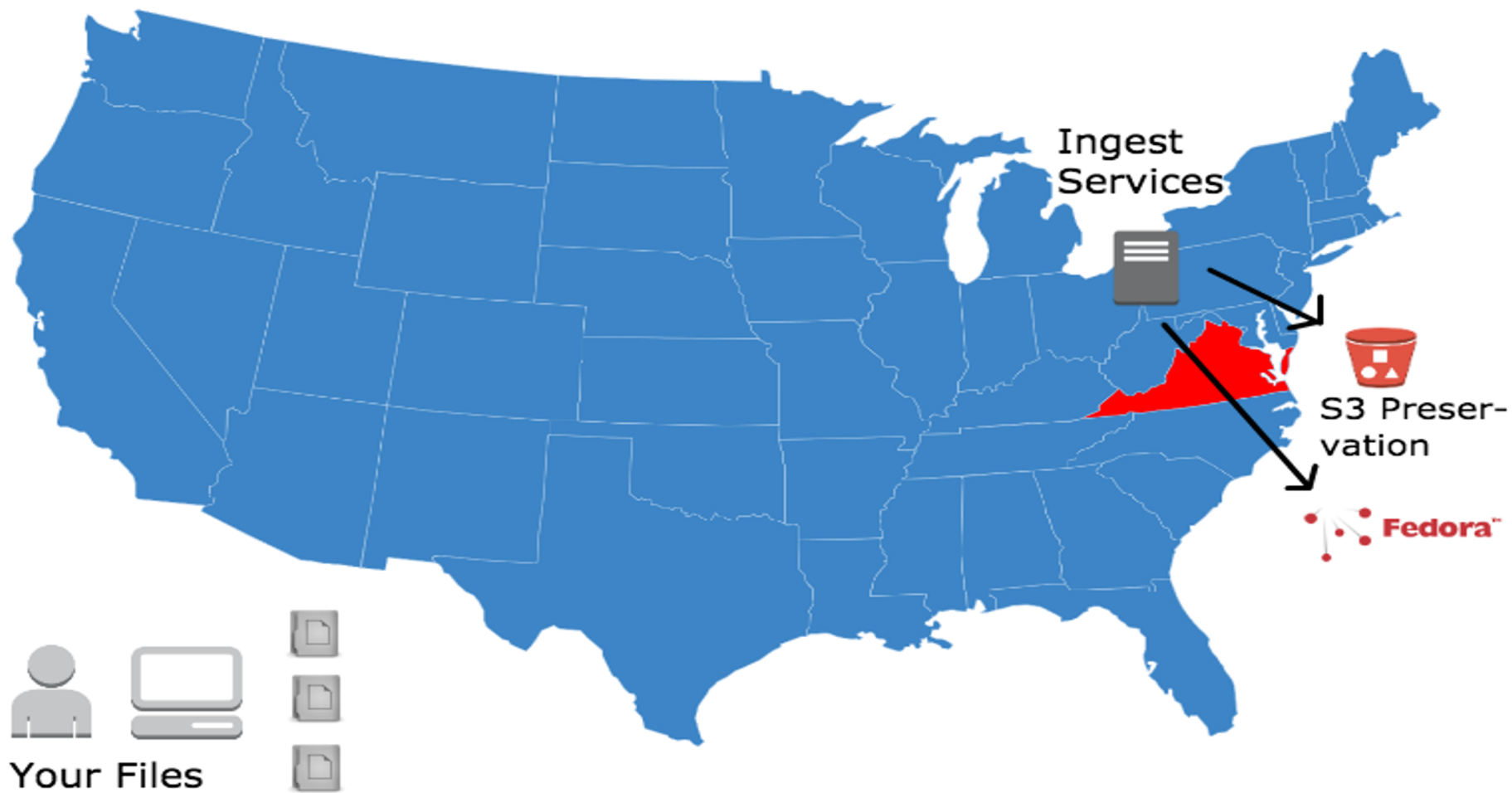
How it works

Andrew Diamond
Senior Preservation & Systems Engineer



Your Files





Glacier



Ingest
Services



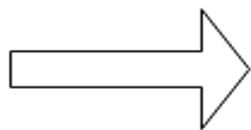


ACADEMIC
PRESERVATION TRUST

Behind the Scenes, Step by Step



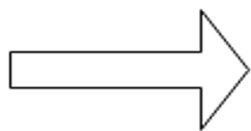
Your Files



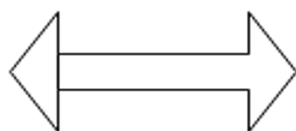
S3
Receiving
(VA)



Your Files



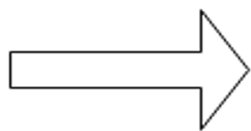
S3
Receiving
(VA)



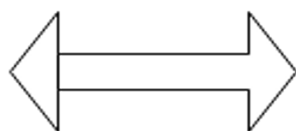
Ingest Server



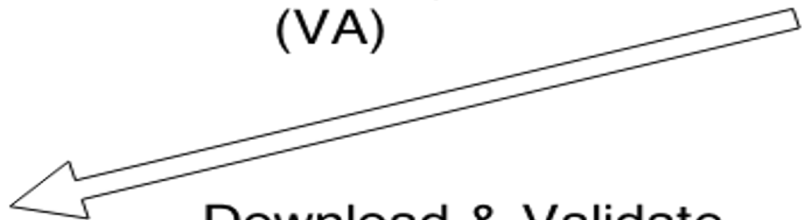
Your Files



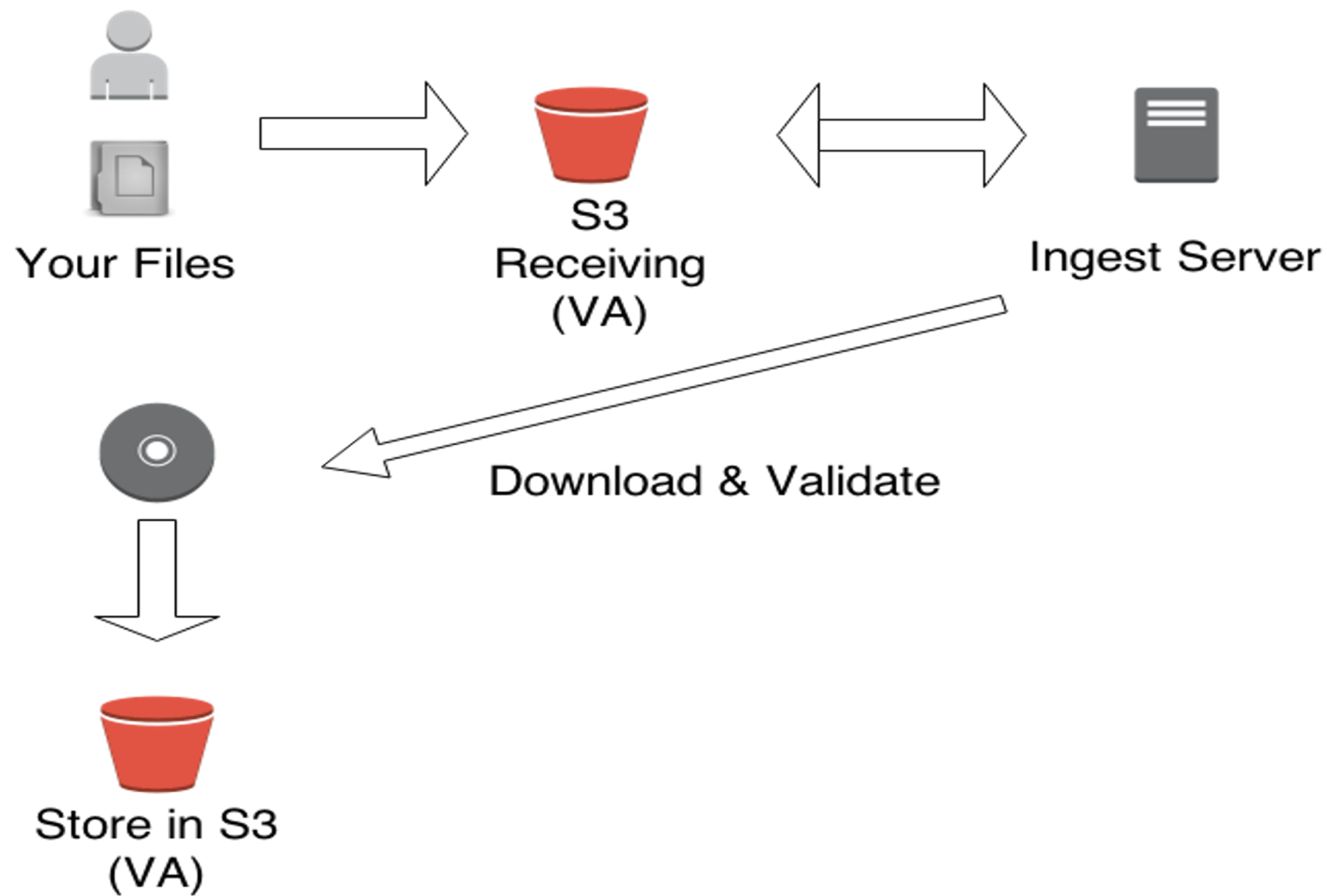
S3
Receiving
(VA)

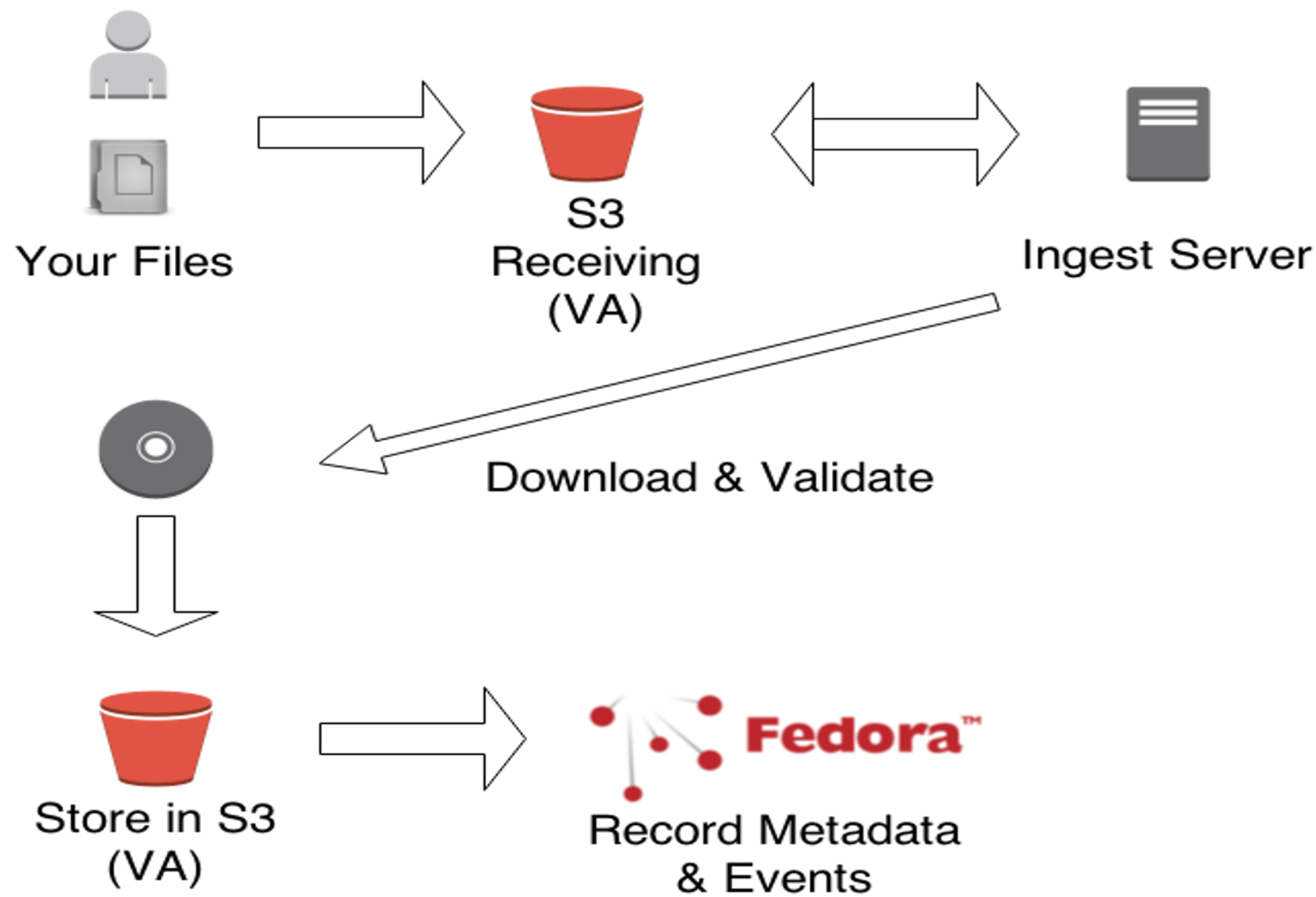


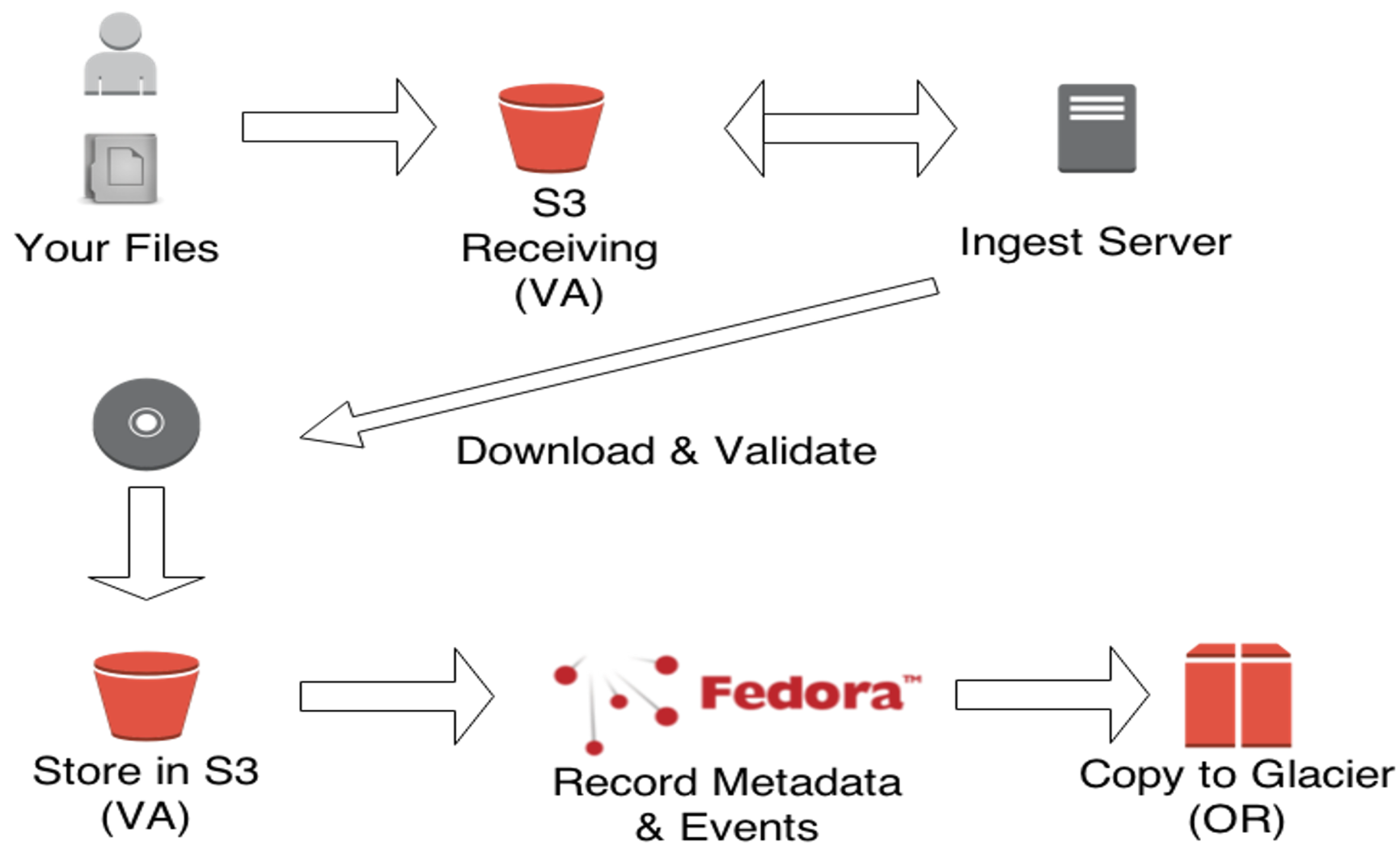
Ingest Server



Download & Validate









ACADEMIC
PRESERVATION TRUST

What You See, Step by Step

Validation Pending

Name	Etag	Institution	Processed	Action	Stage	Status	Reviewed?
virginia.edu.uva-lib_2278801.tar	da04d76f9dca11455297827426db35f2	test.edu	1 minute	Ingest	Validate	Pending	False
<input type="checkbox"/> virginia.edu.uva-lib_2274642.tar	01bf0136e50c6c475bdc31f0323f3310	test.edu	1 minute	Ingest	Validate	Failed	False
virginia.edu.uva-lib_2274765.tar	3f43304f3f7c8d51111d0846e13cb74e	test.edu	1 minute	Ingest	Validate	Pending	False
ncsu.1840.16-2928.tar	b4f8f3072f73598fc5b65bf416b6019a	test.edu	1 minute	Ingest	Validate	Pending	False
virginia.edu.uva-lib_2141114.tar	6e46b8a0e17a968aa8d88d279d50ed44	test.edu	1 minute	Ingest	Validate	Pending	False
ncsu.1840.16-1005.tar	bd350f3229d426fd23fd3fb64bc5e576	test.edu	1 minute	Ingest	Validate	Pending	False
ncsu.1840.16-1028.tar	1e6f9764fed98ae83453bb6d0b679cc0	test.edu	1 minute	Ingest	Validate	Pending	False
ncsu.1840.16-1013.tar	1f0c214d33c30dba20c98ee852991600	test.edu	1 minute	Ingest	Validate	Pending	False
ncsu.1840.16-10.tar	845f9be9c825b668f3bae1d23ceb01de	test.edu	1 minute	Ingest	Validate	Pending	False
ncsu.1840.16-1004.tar	3d4fede4b748a5d8acbdefd13f39d0cd	test.edu	1 minute	Ingest	Validate	Pending	False

Validation Succeeded (Detail)

virginia.edu.uva-lib_2278801.tar

Name	virginia.edu.uva-lib_2278801.tar
Institution	test.edu
Bucket	aptrust.receiving.test.test.edu
Etag	da04d76f9dca11455297827426db35f2
Object Identifier	
Bag Date	2015-01-27 15:30:45 UTC
Date Processed	2015-01-27 15:34:59 UTC
Action	Ingest
Stage	Validate
Status	Pending
Note / Error	No problems

Validation Failed (Detail)

virginia.edu.uva-lib_2274642.tar

Name	virginia.edu.uva-lib_2274642.tar
Institution	test.edu
Bucket	aptrust.receiving.test.test.edu
Etag	01bf0136e50c6c475bdc31f0323f3310
Object Identifier	
Bag Date	2015-01-27 15:30:45 UTC
Date Processed	2015-01-27 15:34:58 UTC
Action	Ingest
Stage	Validate
Status	Failed
Note / Error	Required field Title is missing from tag file.

Storage Pending

Name	Etag	Institution	Processed	Action	Stage	Status	Reviewed?
virginia.edu.uva-lib_2278801.tar	da04d76f9dca11455297827426db35f2	test.edu	less than a minute	Ingest	Store	Pending	False
ncsu.1840.16-1013.tar	1f0c214d33c30dba20c98ee852991600	test.edu	less than a minute	Ingest	Store	Pending	False
ncsu.1840.16-1028.tar	1e6f9764fed98ae83453bb6d0b679cc0	test.edu	less than a minute	Ingest	Store	Pending	False
virginia.edu.uva-lib_2274765.tar	3f43304f3f7c8d51111d0846e13cb74e	test.edu	less than a minute	Ingest	Store	Pending	False
ncsu.1840.16-2928.tar	b4f8f3072f73598fc5b65bf416b6019a	test.edu	less than a minute	Ingest	Store	Pending	False
ncsu.1840.16-1005.tar	bd350f3229d426fd23fd3fb64bc5e576	test.edu	less than a minute	Ingest	Store	Pending	False
ncsu.1840.16-10.tar	845f9be9c825b668f3bae1d23ceb01de	test.edu	less than a minute	Ingest	Store	Pending	False
ncsu.1840.16-1004.tar	3d4fede4b748a5d8acbdefd13f39d0cd	test.edu	less than a minute	Ingest	Store	Pending	False
virginia.edu.uva-lib_2141114.tar	6e46b8a0e17a968aa8d88d279d50ed44	test.edu	less than a minute	Ingest	Store	Pending	False

Record Pending

Name	Etag	Institution	Processed	Action	Stage	Status	Reviewed?
ncsu.1840.16-1005.tar	bd350f3229d426fd23fd3fb64bc5e576	test.edu	1 minute	Ingest	Record	Started	False
ncsu.1840.16-10.tar	845f9be9c825b668f3bae1d23ceb01de	test.edu	1 minute	Ingest	Record	Started	False
ncsu.1840.16-1004.tar	3d4fede4b748a5d8acbdefd13f39d0cd	test.edu	1 minute	Ingest	Record	Started	False
virginia.edu.uva-lib_2141114.tar	6e46b8a0e17a968aa8d88d279d50ed44	test.edu	1 minute	Ingest	Record	Started	False
virginia.edu.uva-lib_2278801.tar	da04d76f9dca11455297827426db35f2	test.edu	1 minute	Ingest	Store	Pending	False
ncsu.1840.16-1028.tar	1e6f9764fed98ae83453bb6d0b679cc0	test.edu	1 minute	Ingest	Store	Pending	False
virginia.edu.uva-lib_2274765.tar	3f43304f3f7c8d51111d0846e13cb74e	test.edu	1 minute	Ingest	Store	Pending	False
ncsu.1840.16-1013.tar	1f0c214d33c30dba20c98ee852991600	test.edu	1 minute	Ingest	Store	Pending	False
ncsu.1840.16-2928.tar	b4f8f3072f73598fc5b65bf416b6019a	test.edu	1 minute	Ingest	Store	Pending	False

Record Succeeded

	Name	Etag	Institution	Processed	Action	Stage	Status	Reviewed?
<input type="checkbox"/>	virginia.edu.uva-lib_2278801.tar	da04d76f9dca11455297827426db35f2	test.edu	3 minutes	Ingest	Record	Success	False
<input type="checkbox"/>	ncsu.1840.16-10.tar	845f9be9c825b668f3bae1d23ceb01de	test.edu	3 minutes	Ingest	Record	Success	False
<input type="checkbox"/>	ncsu.1840.16-1028.tar	1e6f9764fed98ae83453bb6d0b679cc0	test.edu	3 minutes	Ingest	Record	Success	False
<input type="checkbox"/>	virginia.edu.uva-lib_2274765.tar	3f43304f3f7c8d51111d0846e13cb74e	test.edu	3 minutes	Ingest	Record	Success	False
<input type="checkbox"/>	ncsu.1840.16-1004.tar	3d4fede4b748a5d8acbddefd13f39d0cd	test.edu	3 minutes	Ingest	Record	Success	False
<input type="checkbox"/>	ncsu.1840.16-1013.tar	1f0c214d33c30dba20c98ee852991600	test.edu	3 minutes	Ingest	Record	Success	False
<input type="checkbox"/>	ncsu.1840.16-2928.tar	b4f8f3072f73598fc5b65bf416b6019a	test.edu	3 minutes	Ingest	Record	Success	False
<input type="checkbox"/>	virginia.edu.uva-lib_2141114.tar	6e46b8a0e17a968aa8d88d279d50ed44	test.edu	3 minutes	Ingest	Record	Success	False
<input type="checkbox"/>	ncsu.1840.16-1005.tar	bd350f3229d426fd23fd3fb64bc5e576	test.edu	3 minutes	Ingest	Record	Success	False

List of Ingested Objects

1 - 8 of 8

Sort by relevance ▾

10 per page ▾

1. [Strabo De situ orbis.](#)

 View

Institution: Test University

Identifier: test.edu/virginia.edu.uva-lib_2278801

Bag Name: virginia.edu.uva-lib_2278801

Last Modified: 2015-04-02T12:59:45Z

2. [Parents mad as Charlottesville schools close over integration](#)

 View

Institution: Test University

Identifier: test.edu/virginia.edu.uva-lib_2274765

Bag Name: virginia.edu.uva-lib_2274765

Last Modified: 2015-04-02T12:58:43Z

3. [Finale til Baletten: Livjægerne paa Amager \[Galop\]](#)

 View

Institution: Test University

Identifier: test.edu/virginia.edu.uva-lib_2141114

Bag Name: virginia.edu.uva-lib_2141114

Last Modified: 2015-04-02T12:58:35Z

Object Detail

Effect of Computer Animation Instruction on Spatial Visualization Performance

Institution [Test University](#) [test]

Access consortia

Identifier test.edu/ncsu.1840.16-1028

Bag Name ncsu.1840.16-1028

Last Modified 2015-04-02T12:58:25.273Z

Description

 [View Preserved Files](#)

 [View Events](#)

 [Restore Object](#)

 [Delete](#)

Aggregate Data About Preserved Files

Files Preserved	4
Total File Size	344 KB
text/plain	1
application/xml	2
application/pdf	1

List of Files

Files for Effect of Computer Animation Instruction on Spatial Visualization Performance


Object Identifier: test.edu/ncsu.1840.16-1028

1 - 4 of 4

10 per page ▾

Sort by relevance ▾

1. <test.edu/ncsu.1840.16-1028/data/ORIGINAL/1-metadata.xml>

 View

Institution: Test University


MimeType: application/xml

Filesize: 128 Bytes

Created: 2014-04-25T13:51:10-04:00

Last Modified: 2014-04-25T13:51:10-04:00

2. <test.edu/ncsu.1840.16-1028/data/ORIGINAL/1>

 View

Institution: Test University

MimeType: application/pdf

Filesize: 339 KB

Created: 2014-04-25T13:51:10-04:00

Last Modified: 2014-04-25T13:51:10-04:00

3. <test.edu/ncsu.1840.16-1028/data/object.properties>

 View

Institution: Test University

File Detail

Test University / Effect of Computer Animation Instruction on Spatial Visualization Performance

/ test.edu/ncsu.1840.16-1028/data/ORIGINAL/1-metadata.xml

test.edu/ncsu.1840.16-1028/data/ORIGINAL/1-metadata.xml

Institution	Test University [test]
Intellectual Object	Effect of Computer Animation Instruction on Spatial Visualization Performance [consortia]
Identifier	test.edu/ncsu.1840.16-1028/data/ORIGINAL/1-metadata.xml
URL to File	https://s3.amazonaws.com/aptrust.test.preservation/059db6bd-1645-46d6-7df7-32ea8d1f70bc
MimeType	application/xml
Filesize	128 Bytes
Created	2014-04-25T13:51:10-04:00 (11 months)
Modified	2014-04-25T13:51:10-04:00 (11 months)
Checksums	md5: a4d9c67041d961bb8e003fdc2a3b65e8 generated 11 months ago sha256: 60eca5faef45a627d0c1e916026ed6cf91ffe911b5f9b136a3dcc7d99e291519 generated about 7 hours ago

 [View Events](#)

 [Delete](#)

Events

1 - 6 of 6

10 per page ▾

1. [test.edu/ncsu.1840.16-1013/data/ORIGINAL/1](#)

Event Identifier: 0488ca41-02b0-4684-78f3-e099e711305e

Event Type: ingest

DateTime: 2015-01-27T16:21:55.986658362Z

Detail: Ingested to replication storage and assigned replication URL identifier

Outcome: Success

Outcome Details: <https://s3.amazonaws.com/aptrust.test.preservation.oregon/c1ee56f2-4897-4a1a-601a-eeff204b94d0>

Object: Go uuid library + goamz S3 library

Agent: <http://github.com/nu7hatch/gouuid>

2. [test.edu/ncsu.1840.16-1013/data/ORIGINAL/1](#)

Event Identifier: be7181ef-4560-42c1-447d-55c4e3fdb7d9

Event Type: fixity_generation

DateTime: 2015-01-27T16:19:13.263192327Z

Detail: Calculated new fixity value

Outcome: Success

Outcome Details: [sha256:5cbf7f3fd8461ff0319249821ad5fd0b378303302611c5264d18b88a3250e18a](#)

Object: Go language crypto/sha256

Agent: <http://golang.org/pkg/crypto/sha256/>

3. [test.edu/ncsu.1840.16-1013/data/ORIGINAL/1](#)

Event Identifier: 115105f7-94de-4441-4528-6b5d3e8da2ef
Event Type: identifier_assignment
DateTime: 2015-01-27T16:19:13.262975864Z
Detail: Assigned new institution.bag/path identifier
Outcome: Success
Outcome Details: test.edu/ncsu.1840.16-1013/data/ORIGINAL/1
Object: APTrust bag processor
Agent: <https://github.com/APTrust/bagman>

4. [test.edu/ncsu.1840.16-1013/data/ORIGINAL/1](#)

Event Identifier: 4d863ca9-8100-4da5-6d21-cd3cedc887ae
Event Type: identifier_assignment
DateTime: 2015-01-27T16:19:13.262975864Z
Detail: Assigned new storage URL identifier
Outcome: Success
Outcome Details: <https://s3.amazonaws.com/aptrust.test.preservation/c1ee56f2-4897-4a1a-601a-eeff204b94d0>
Object: Go uuid library + goamz S3 library
Agent: <http://github.com/nu7hatch/gouuid>

Your item has been queued for restoration.

[Test University](#) / Multicasting in a Partially Tunable Broadcast WDM Network

Multicasting in a Partially Tunable Broadcast WDM Network

Institution [Test University](#) [test]

Access consortia

Identifier test.edu/ncsu.1840.16-1013

Bag Name ncsu.1840.16-1013

Last Modified 2015-01-27T16:21:23.816Z

Description

 [View Preserved Files](#)

 [View Events](#)

 [Restore Object](#)

 [Edit](#)

 [Delete](#)

Pending Tasks: Restore & Delete

Name	Etag	Institution	Processed	Action	Stage	Status	Reviewed?
ncsu.1840.16-1013.tar	1f0c214d33c30dba20c98ee852991600	test.edu	1 minute	Restore	Requested	Pending	False
virginia.edu.uva-lib_2278801.tar	da04d76f9dca11455297827426db35f2	test.edu	2 minutes	Delete	Requested	Pending	False
virginia.edu.uva-lib_2278801.tar	da04d76f9dca11455297827426db35f2	test.edu	2 minutes	Delete	Requested	Pending	False
virginia.edu.uva-lib_2278801.tar	da04d76f9dca11455297827426db35f2	test.edu	2 minutes	Delete	Requested	Pending	False
virginia.edu.uva-lib_2278801.tar	da04d76f9dca11455297827426db35f2	test.edu	2 minutes	Delete	Requested	Pending	False
<input type="checkbox"/> virginia.edu.uva-lib_2278801.tar	da04d76f9dca11455297827426db35f2	test.edu	4 minutes	Ingest	Record	Success	False
<input type="checkbox"/> ncsu.1840.16-1013.tar	1f0c214d33c30dba20c98ee852991600	test.edu	4 minutes	Ingest	Record	Success	False

Completed Tasks: Restore & Delete

	Name	Etag	Institution	Processed	Action	Stage	Status	Reviewed?
<input type="checkbox"/>	ncsu.1840.16-1013.tar	1f0c214d33c30dba20c98ee852991600	test.edu	3 minutes	Restore	Resolve	Success	False
<input type="checkbox"/>	virginia.edu.uva-lib_2278801.tar	da04d76f9dca11455297827426db35f2	test.edu	4 minutes	Delete	Resolve	Success	False
<input type="checkbox"/>	virginia.edu.uva-lib_2278801.tar	da04d76f9dca11455297827426db35f2	test.edu	4 minutes	Delete	Resolve	Success	False
<input type="checkbox"/>	virginia.edu.uva-lib_2278801.tar	da04d76f9dca11455297827426db35f2	test.edu	4 minutes	Delete	Resolve	Success	False
<input type="checkbox"/>	virginia.edu.uva-lib_2278801.tar	da04d76f9dca11455297827426db35f2	test.edu	4 minutes	Delete	Resolve	Success	False
<input type="checkbox"/>	virginia.edu.uva-lib_2278801.tar	da04d76f9dca11455297827426db35f2	test.edu	5 minutes	Ingest	Record	Success	False
<input type="checkbox"/>	ncsu.1840.16-1013.tar	1f0c214d33c30dba20c98ee852991600	test.edu	6 minutes	Ingest	Record	Success	False

Detail of Completed Restore

ncsu.1840.16-1013.tar

Name	ncsu.1840.16-1013.tar
Institution	test.edu
Bucket	aptrust.receiving.test.test.edu
Etag	1f0c214d33c30dba20c98ee852991600
Object Identifier	test.edu/ncsu.1840.16-1013
Bag Date	2015-01-27 15:30:44 UTC
Date Processed	2015-01-27 16:24:51 UTC
Action	Restore
Stage	Resolve
Status	Success
Note / Error	Object restored to https://s3.amazonaws.com/aptrust.test.restore/ncsu.1840.16-1013.tar

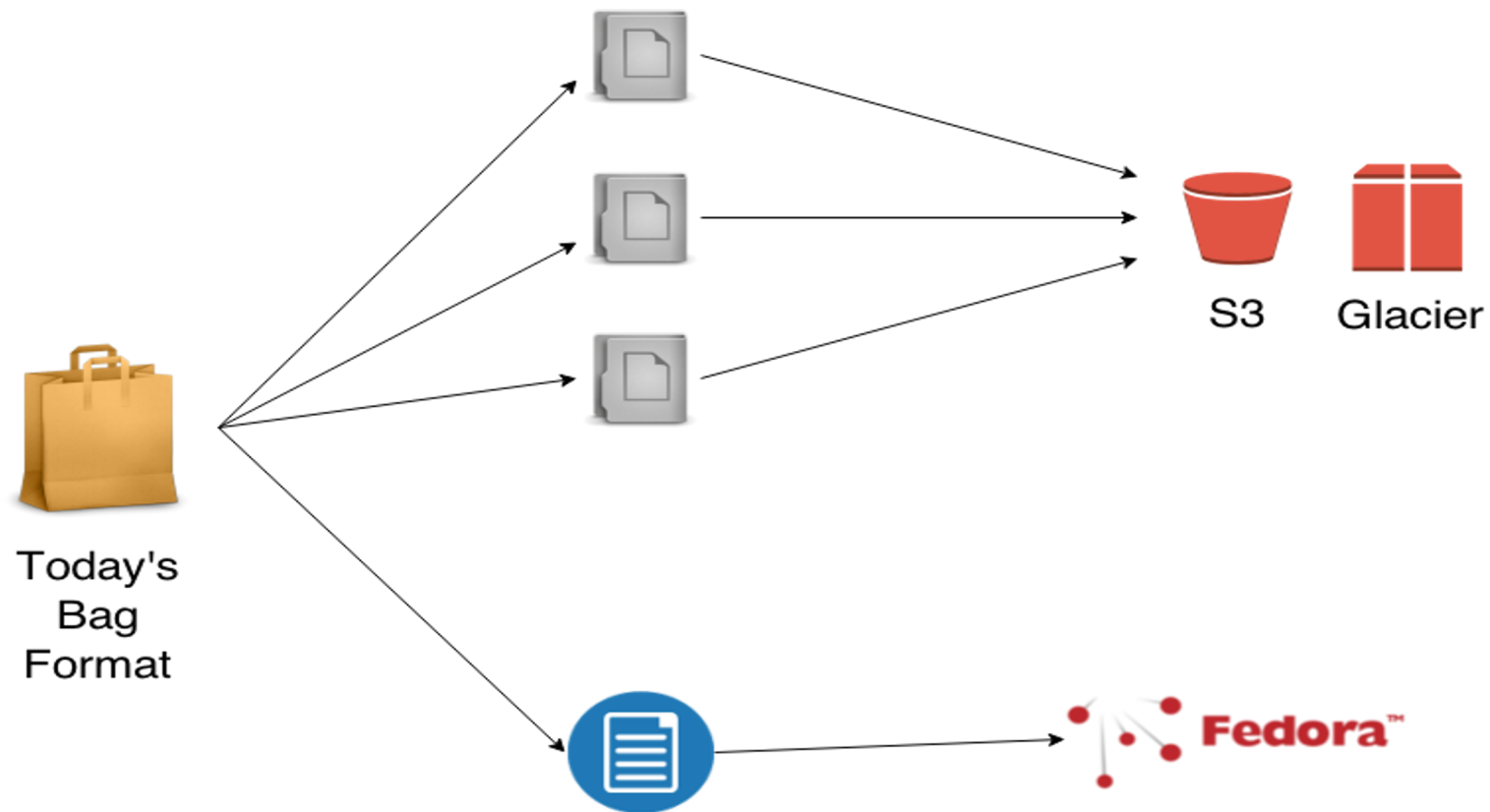
Detail of Completed Delete

virginia.edu.uva-lib_2278801.tar

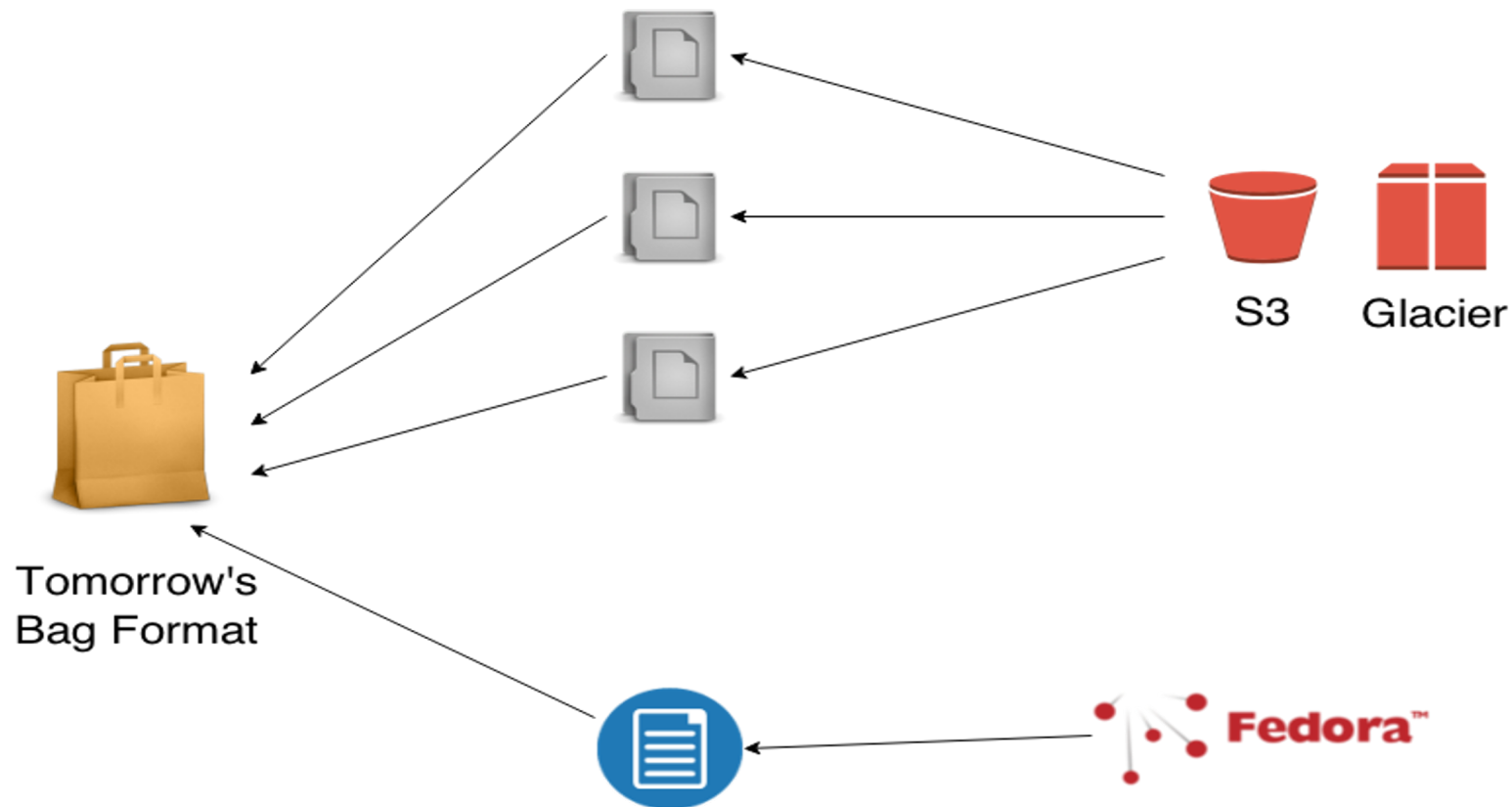
Name	virginia.edu.uva-lib_2278801.tar
Institution	test.edu
Bucket	aptrust.receiving.test.test.edu
Etag	da04d76f9dca11455297827426db35f2
Object Identifier	test.edu/virginia.edu.uva-lib_2278801
Generic File Identifier	test.edu/virginia.edu.uva-lib_2278801/data/datastream-descMetadata
Bag Date	2015-01-27 15:30:45 UTC
Date Processed	2015-01-27 16:23:35 UTC
Action	Delete
Stage	Resolve
Status	Success
Note / Error	Deleted generic file 'test.edu/virginia.edu.uva-lib_2278801/data/datastream-descMetadata' from 'https://s3.amazonaws.com/aptrust.test.preservation/bdb83294-1d02-4b76-4aad-a207e7c1fe47' at 2015-01-27T11:27:36-05:00 at the request of andrew.diamond@aptrust.org

A Note on Bag Storage & Restoration

Bag Storage



Bag Restoration





ACADEMIC
PRESERVATION TRUST

Where's my data?



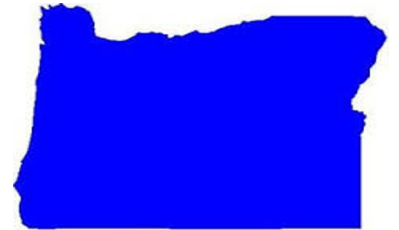
S3

On multiple disks in
Northern Virginia.



Glacier

On hard disk, tape or
optical disk in Oregon.





S3 = Simple Storage Service



S3 = Simple Storage Service

99.999999999% Durability (9 nines)



S3 = Simple Storage Service

99.999999999% Durability (9 nines)

Each file stored on multiple devices in
multiple data centers



S3 = Simple Storage Service

99.999999999% Durability (9 nines)

Each file stored on multiple devices in
multiple data centers

Fixity check on transfer



S3 = Simple Storage Service

99.999999999% Durability (9 nines)

Each file stored on multiple devices in multiple data centers

Fixity check on transfer

Ongoing data integrity checks & self-healing



Glacier

Long-term storage with 99.9999999999% durability (11 nines)

Not intended for frequent access

Takes five hours or so to restore a file



ACADEMIC
PRESERVATION TRUST

Chances of losing an archived object?

Chances of losing an archived object?



1 in 100,000,000,000

S3

Chances of losing an archived object?



1 in 100,000,000,000

S3



1 in 10,000,000,000,000

Glacier

Chances of losing an archived object?



1 in 100,000,000,000

S3



1 in 10,000,000,000,000

Glacier

1 in 1,000,000,000,000,000,000,000,000,000
chance of losing both to separate events

Chances of losing an archived object?



1 in 100,000,000,000

S3



1 in 10,000,000,000,000

Glacier

Very Slim

~~1 in 1,000,000,000,000,000,000,000,000~~

chance of losing both to separate events



ACADEMIC
PRESERVATION TRUST

Protecting Against Failure During Ingest



ACADEMIC
PRESERVATION TRUST

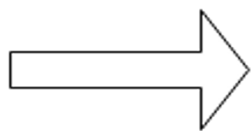
Protecting Against Failure During Ingest

Microservices with state

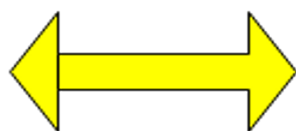
Resume intelligently after failure



Your Files



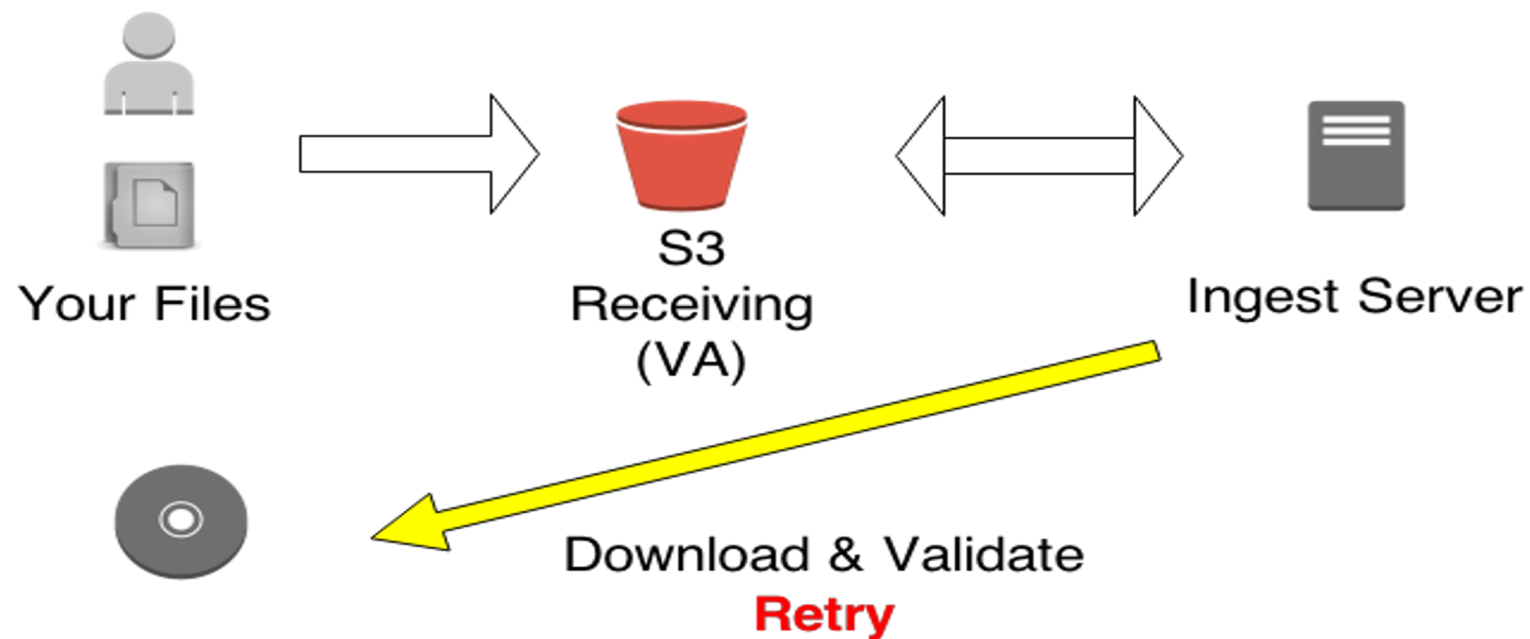
S3
Receiving
(VA)

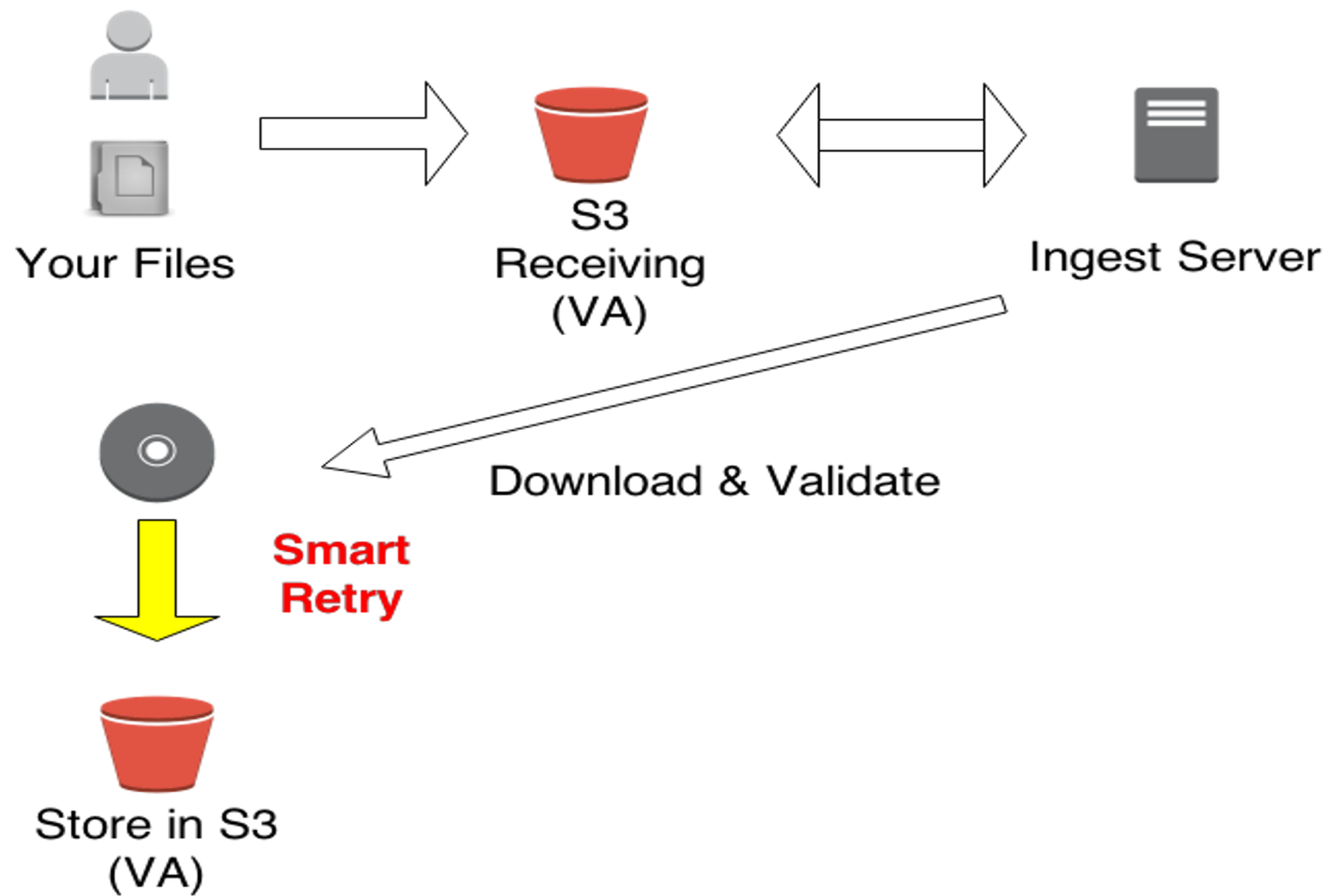


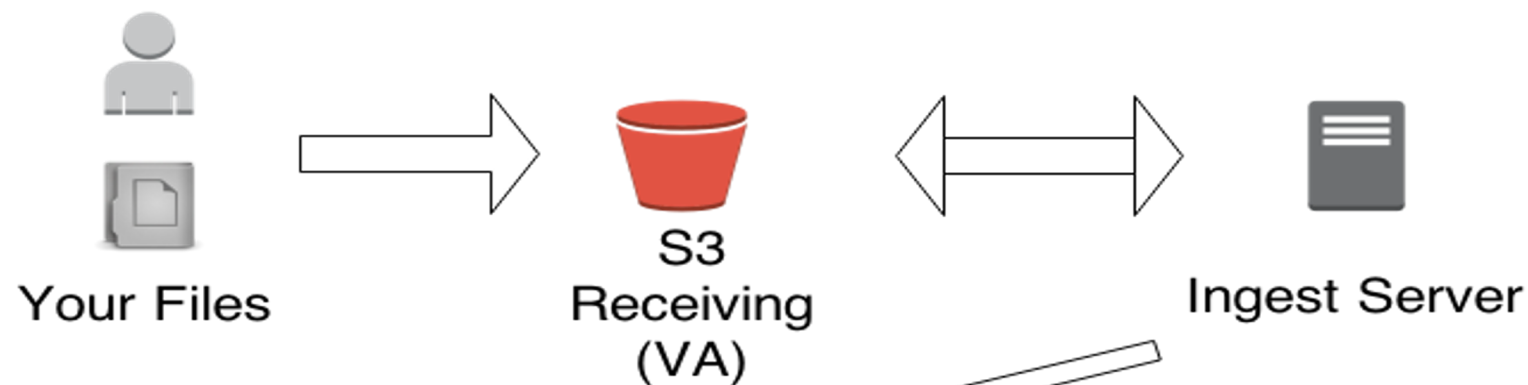
Retry



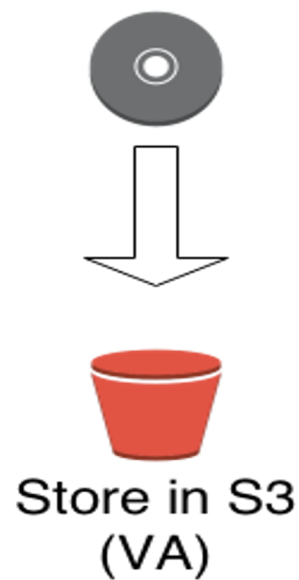
Ingest Server





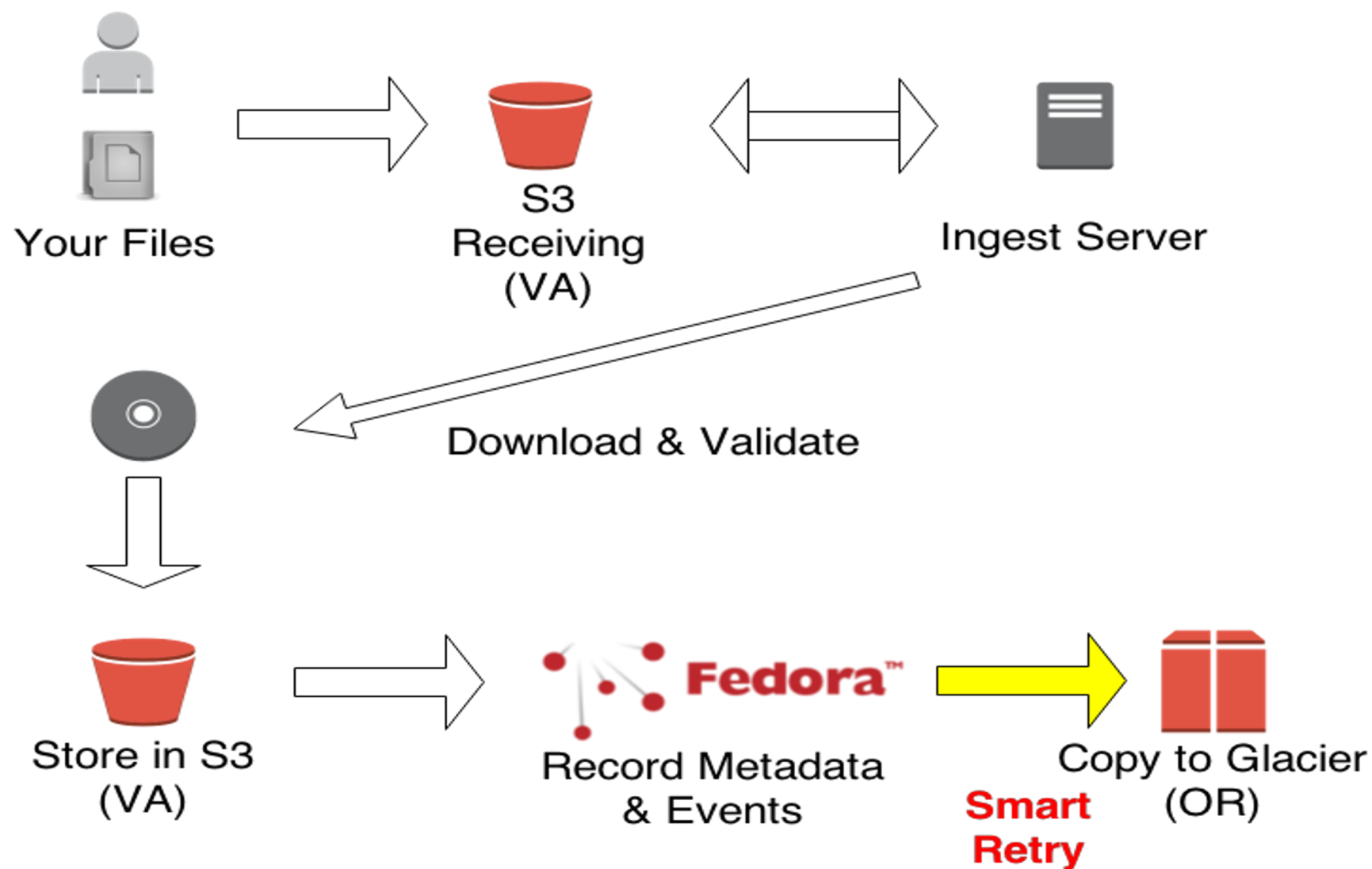


Download & Validate



**Smart
Retry**







ACADEMIC
PRESERVATION TRUST

Protecting Against Failure After Ingest



ACADEMIC
PRESERVATION TRUST

Protecting Against Failure After Ingest



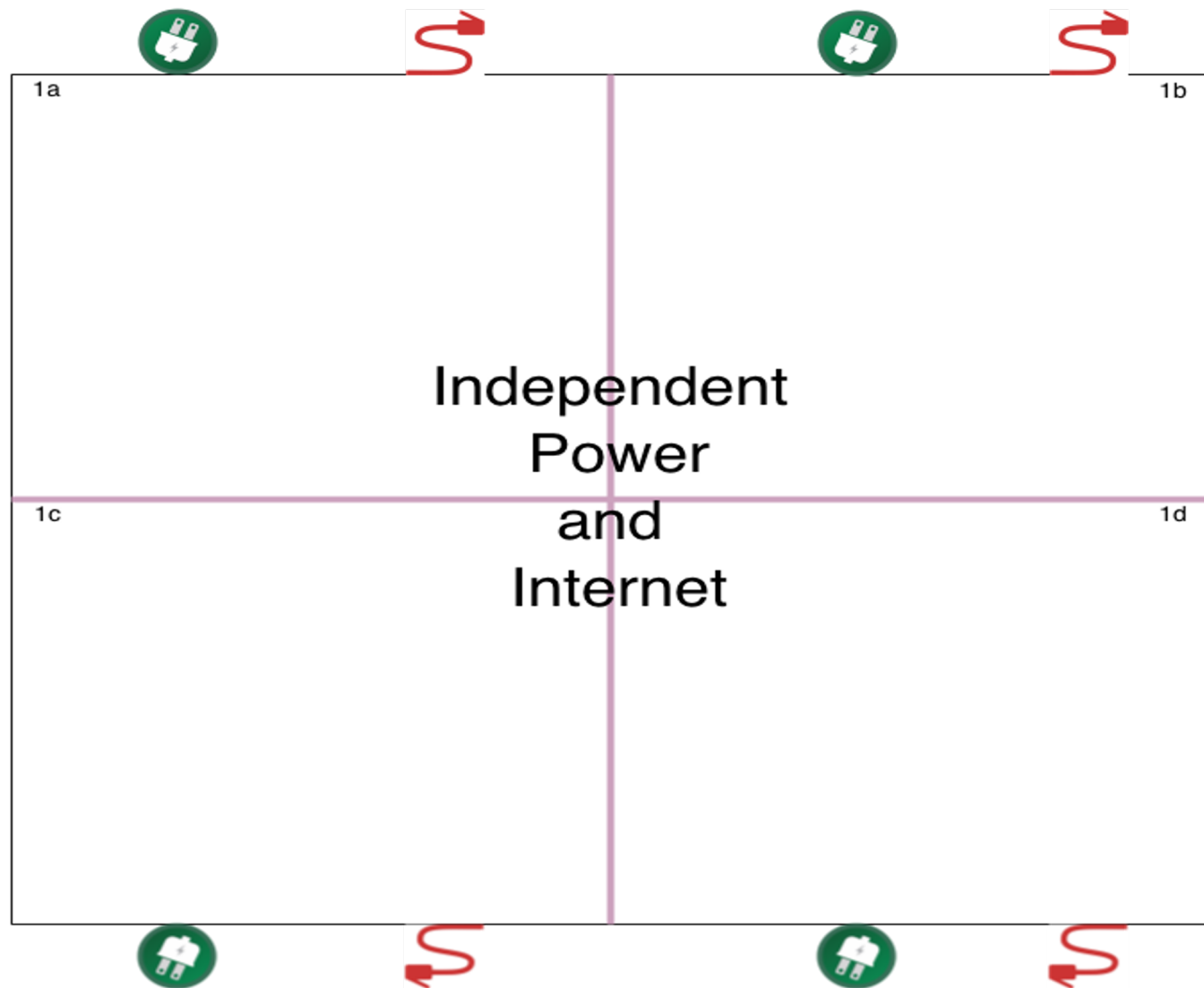
Objects & Files



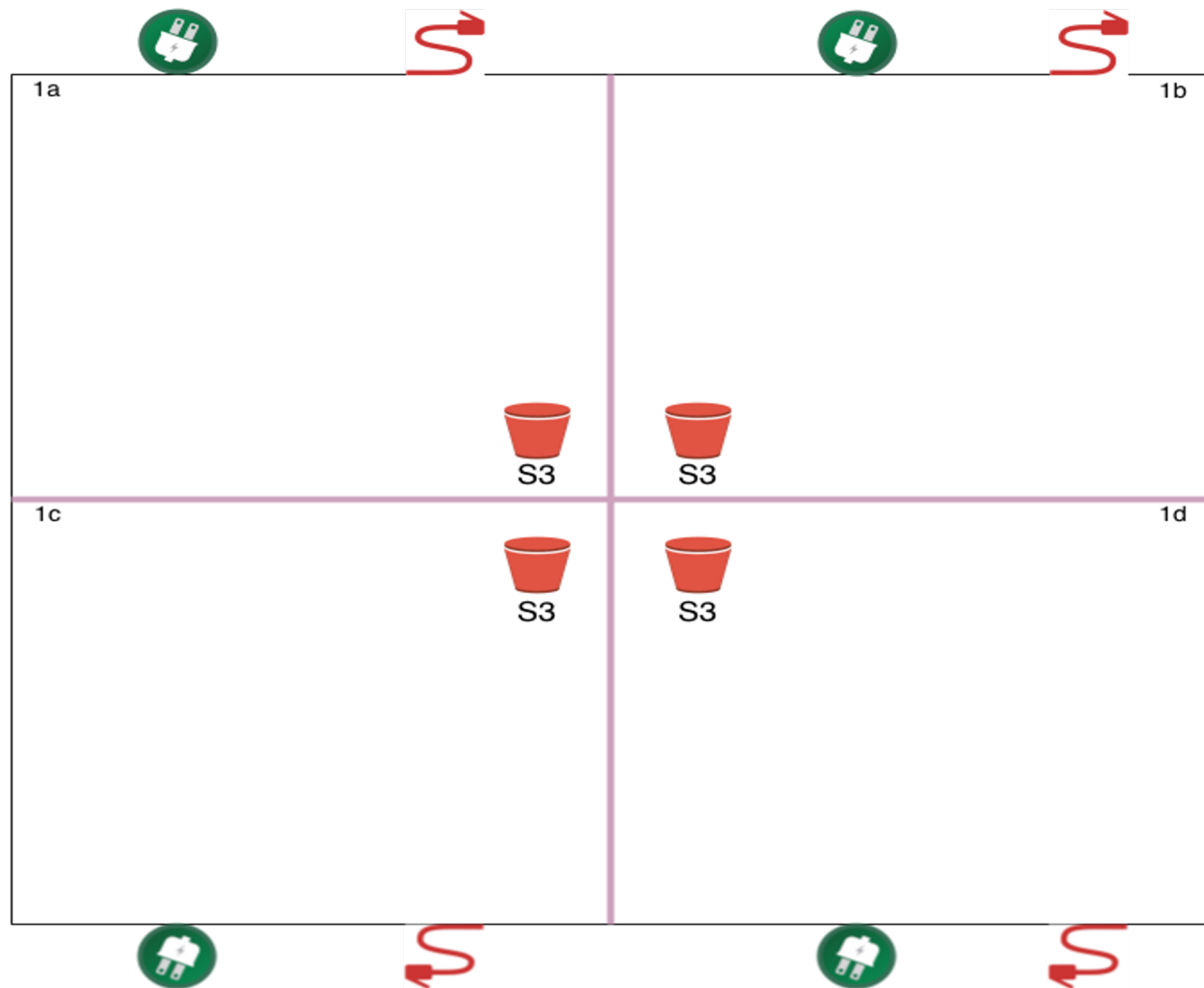
FedoraTM

Metadata & Events

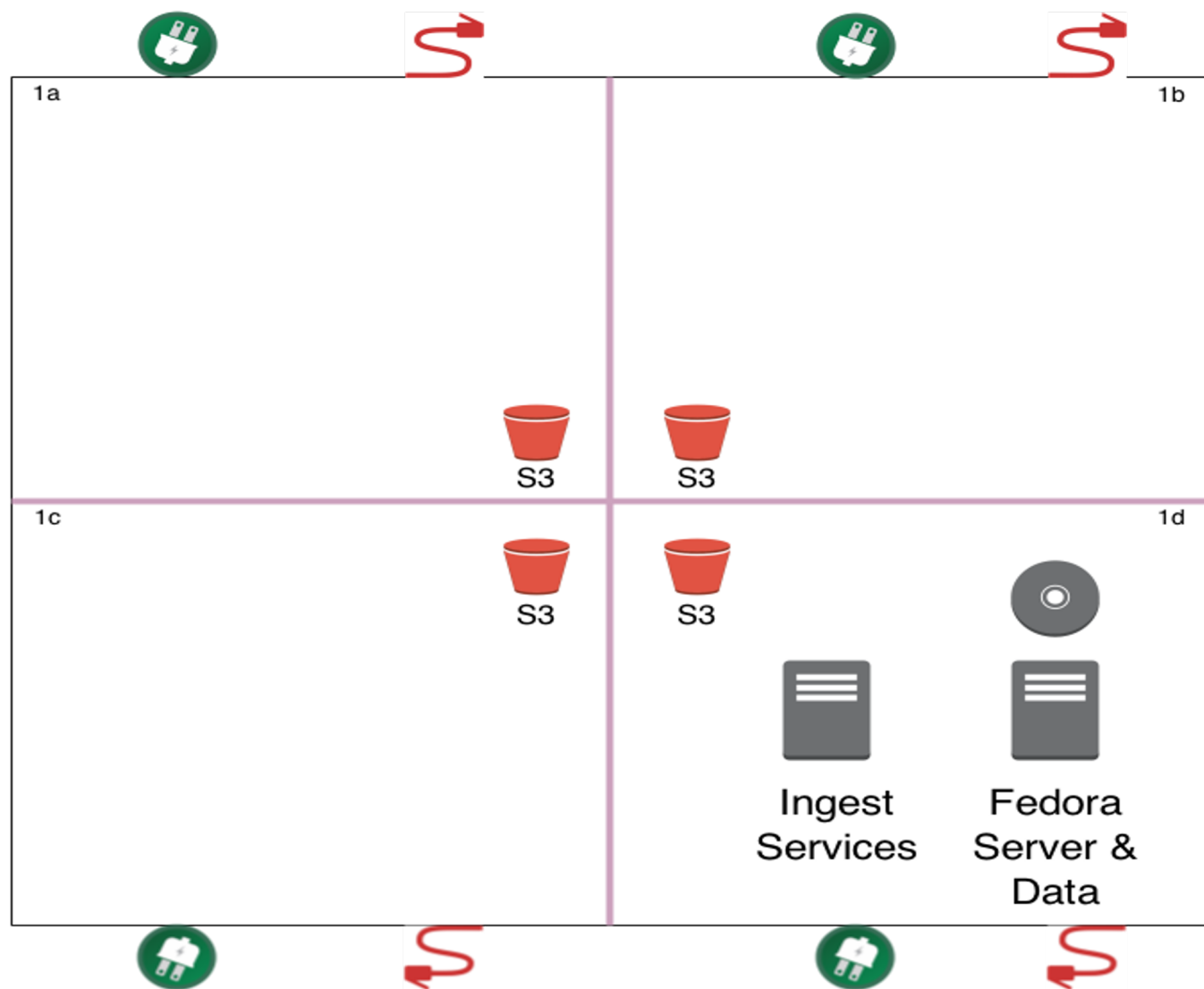
AWS Availability Zones - US East (VA)



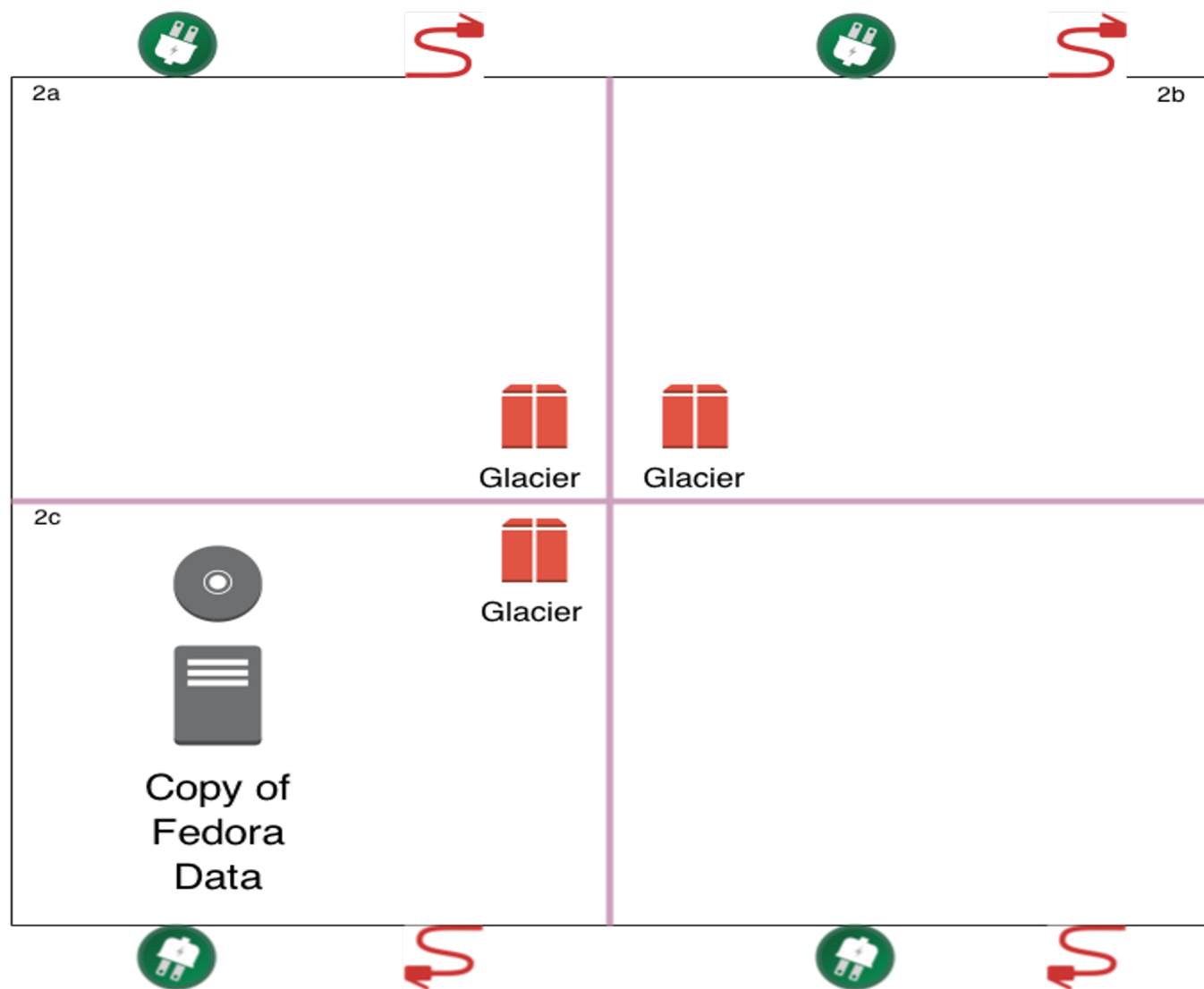
AWS Availability Zones - US East (VA)



AWS Availability Zones - US East (VA)



AWS Availability Zones - US West (OR)



What if...



ACADEMIC
PRESERVATION TRUST

What if ingest services stop?



They restart themselves and pick up where they left off.

What if ingest server dies?



We spin up a new one in a few hours.

What if Fedora service stops?



We restart it manually.

What if Fedora server dies?



We can build a new one from the data on the hot backup in Oregon.

What if we lose all of Glacier in Oregon?



We have all files in S3 in Virginia. Ingest services and Fedora are not affected.

What if we lose the whole Virginia data center?

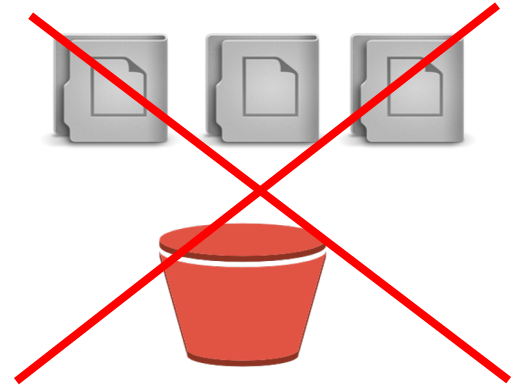


Fedora

**Metadata and
Events**



Ingest
Services



S3

**Files and
Objects**

What if we lose the whole Virginia data center?



Ingest
Services

We can
rebuild
services from
code and
config files in
hours.

What if we lose the whole Virginia data center?

We can restore all files
from Glacier in Oregon.



S3

**Files and
Objects**

What if we lose the whole Virginia data center?




Fedora

**Metadata and
Events**

We can restore all
metadata and
events from the
Fedora backup in
Oregon.

Metadata stored with files

 **AWS** ▾ **Services** ▾ **Edit** ▾

andrew.diamond @ aptrust ▾ Global ▾ Support ▾

Upload **Create Folder** **Actions** ▾

All Buckets / **aptrust.preservation.storage**

	Name	Storage Class	Size	Last Modified
<input type="checkbox"/>	000159e2-a8c6-4cbe-46fb-a6124...	Standard	2.9 KB	Fri Feb 20 16:45:54 G
<input type="checkbox"/>	0001d2f3-9c73-4b87-7c3a-0a642f...	Standard	249 KB	Wed Feb 25 00:00:03
<input type="checkbox"/>	000241fa-03a5-4238-4bdf-28dc3e...	Standard	3 KB	Thu Feb 26 05:59:46
<input type="checkbox"/>	00028c0c-9017-4975-679f-dce34...	Standard	187.3 KB	Wed Feb 25 06:41:33
<input type="checkbox"/>	0003ee14-cca0-46fb-756f-04fe8fc...	Standard	16.9 MB	Sat Feb 21 17:13:30 C
<input type="checkbox"/>	000440be-1888-4eac-7c0b-d9637...	Standard	3.4 KB	Wed Feb 25 22:59:06
<input type="checkbox"/>	000499af-0ce0-460a-4c83-724ed...	Standard	269 bytes	Sat Feb 21 08:12:20 C
<input type="checkbox"/>	0004ce26-586c-4a52-44b4-dae3d...	Standard	2.8 KB	Sat Feb 21 17:11:51 C
<input type="checkbox"/>	00053343-e649-4b26-60ae-60033...	Standard	548 bytes	Wed Feb 25 14:56:00
<input type="checkbox"/>	000550f0-32eb-4292-56ac-ff437c...	Standard	3.6 KB	Sat Feb 21 11:47:48 C
<input checked="" type="checkbox"/>	0005c5ca-a810-485e-52c2-552c8...	Standard	3.8 KB	Wed Feb 25 10:16:31
<input type="checkbox"/>	00064b4f-0bb2-49da-630e-e6e57...	Standard	8.8 MB	Wed Feb 25 09:10:09
<input type="checkbox"/>	0006b23a-f124-49ba-7078-6eef7a...	Standard	15 bytes	Sun Feb 22 13:03:54
<input type="checkbox"/>	0007f726-3de1-4daf-68fe-d9f2411...	Standard	383.7 KB	Tue Feb 24 15:12:04
<input type="checkbox"/>	00087136-0df5-44b0-43bb-a326e...	Standard	269 bytes	Tue Feb 24 06:33:57
<input type="checkbox"/>	0009e08d-9ef9-4db7-69ef-6cb226...	Standard	3.1 KB	Fri Feb 20 11:47:45 G
<input type="checkbox"/>	000b7a37-f33f-4b24-6993-46d82c...	Standard	17.2 MB	Mon Feb 23 07:12:23
<input type="checkbox"/>	000c2ad3-a2d8-40e0-77e8-c1a73...	Standard	3.1 KB	Fri Feb 20 13:23:5 G
<input type="checkbox"/>	000c9f4c-f8ca-4c97-75e7-1b81cc...	Standard	282.8 KB	Fri Feb 20 17:45:40 G
<input type="checkbox"/>	000cd248-00fe-4eef-67c9-2e372b...	Standard	285.8 KB	Wed Feb 25 11:47:41
<input type="checkbox"/>	000dfab6-b480-4601-7016-d6537...	Standard	231.8 KB	Sun Feb 22 11:18:28
<input type="checkbox"/>	000edadf-ff7c-4a00-4a0f-ab32b76...	Standard	10.6 MB	Fri Feb 20 23:17:49 G
<input type="checkbox"/>	000eef0d-95a0-4016-753e-69187...	Standard	3.5 KB	Sat Feb 21 14:26:07 C
<input type="checkbox"/>	00105ed9-1776-413a-5c76-65027...	Standard	29.3 MB	Sat Feb 21 11:42:12 C

Object: 0005c5ca-a810-485e-52c2-552c8e38... ×

Bucket: aptrust.preservation.storage
Name: 0005c5ca-a810-485e-52c2-552c8e38cb78
Link: <https://s3.amazonaws.com/aptrust.preservation.storage/0005c5ca-a810-485e-52c2-552c8e38cb78>
Size: 3935
Last Modified: Wed Feb 25 10:16:31 GMT-500 2015
Owner: aws+aptrust
Etag: 7079fba85a4f46315560a11628aa29f
Expiry Date: None
Expiration Rule: N/A

Details

Permissions

Metadata

Key: x-amz-meta-sha256 Value: 5b4e8687ca13393d9c160e68197 ×


Key: x-amz-meta-institution Value: uc.edu ×


Key: x-amz-meta-md5 Value: 7079fba85a4f46315560a11628aa29f ×

Key: x-amz-meta-bagpath Value: data/1431/dublin_core.xml ×

Key: x-amz-meta-bag Value: cin.702760 ×

Key: Content-Type Value: application/xml ×

 **Add more metadata**

 **Remove selected metadata**

Attached metadata for last-ditch recovery

Object: 0005c5ca-a810-485e-52c2-552c8e38... x

Bucket: aptrust.preservation.storage
Name: 0005c5ca-a810-485e-52c2-552c8e38cb78
Link: <https://s3.amazonaws.com/aptrust.preservation.storage/0005c5ca-a810-485e-52c2-552c8e38cb78>
Size: 3935
Last Modified: Wed Feb 25 10:16:31 GMT-500 2015
Owner: aws+aptrust
ETag: 7079fbba85a4f46315560a11628aa29f
Expiry Date: None
Expiration Rule: N/A

▸ Details

▸ Permissions

▾ Metadata

Key: x-amz-meta-sha256	Value: 5b4e8687ca13393d9c160e68197	x
Key: x-amz-meta-institution	Value: uc.edu	x
Key: x-amz-meta-md5	Value: 7079fbba85a4f46315560a11628i	x
Key: x-amz-meta-bagpath	Value: data/1431/dublin_core.xml	x
Key: x-amz-meta-bag	Value: cin.702760	x
Key: Content-Type	Value: application/xml	x

To lose everything...

To lose everything...

A major catastrophic event in
Virginia and Oregon, or...

To lose everything...

A major catastrophic event in
Virginia and Oregon, or...

Don't pay the AWS bill

Partner Tools



ACADEMIC
PRESERVATION TRUST

Partner Tools



ACADEMIC
PRESERVATION TRUST

Bag Management

Upload

Download

List

Delete

Partner Tools



ACADEMIC
PRESERVATION TRUST

Bag Management

Upload

Download

List

Delete

...or use open-source AWS libraries for Java, Python, Ruby,
etc.

Partner Tools



ACADEMIC
PRESERVATION TRUST

Bag Management

1. bash

```
d-128-193-157:bin apd4n$ apt_list --bucket=receiving
Using default config file /Users/apd4n/.aptrust_partner.conf
Listing up to 100 items from bucket aptrust.receiving.testbucket.edu
```

LastModified	ETag	Size	File
2014-04-25T15:31:01.000Z	5b00fb5397e81ce6b231562db67a6729	257250	billmurray_friday.png
2014-12-04T19:09:26.000Z	8f94faf02784d1dee63ce1baef43d998	567	cleanup_result.json
2014-12-08T21:10:09.000Z	b46f7f8588c10a3242612b4f19395d65	16383488	example.edu.multi_mb_test_bag.tar
2015-01-27T16:16:48.000Z	b4c15fd77864df078bf422dc010414eb	24064	example.edu.multipart.b01.of02.tar
2015-01-27T16:16:48.000Z	accb01c0ba1561420d1b576323096d32	24064	example.edu.multipart.b02.of02.tar
2014-12-08T21:10:13.000Z	4a7bdc37096ca95d3820159759a38bac	47	example.edu.not_a_tar_file.tar
2014-12-08T21:10:14.000Z	a6e5d86d6c35b28e273de9ac5cba940a	24576	example.edu.sample_bad_access.tar
2014-12-08T21:10:14.000Z	d158308417df8b4d85cb4bf989a5fa53	23552	example.edu.sample_bad_checksums.tar
2014-12-08T21:10:16.000Z	ee98f9e5b64d621ab8fde252591f5822	20480	example.edu.sample_missing_data_file.tar
2014-12-08T21:10:17.000Z	e8da2a06ed8beb07e07f767b9ace492c	29696	example.edu.sample_no_aptrust_info.tar
2014-12-08T21:10:18.000Z	d40da4365398c839e9a9d1c50c53de64	22528	example.edu.sample_no_bag_info.tar
2014-12-08T21:10:19.000Z	f06ae3cd7ec9225cbcade79c2470821c	22528	example.edu.sample_no_bagit.tar
2014-12-08T21:10:20.000Z	c996f2bb15560920797c7ea68120817d	12800	example.edu.sample_no_data_dir.tar
2014-12-08T21:10:21.000Z	41075cf5d423a82304dba948cfef5918	29696	example.edu.sample_no_md5_manifest.tar
2014-12-08T21:10:21.000Z	a2e6b490b98822901e8edc50827191fc	24576	example.edu.sample_no_title.tar
2014-12-08T21:10:22.000Z	3b0aaffe71d2bb765ab435a5d75506c1	23552	example.edu.sample_wrong_folder_name.tar
2014-12-04T19:09:30.000Z	73554782a26b873f830aafd9dd169da4	15927	intel_obj.json
2014-12-04T19:09:30.000Z	c223b8b91a4803d3fdf171dfd6ca5118	277	load_env_test.txt
2014-12-10T21:36:21.000Z	b4f8f3072f73598fc5b65bf416b6019a	696320	ncsu.1840.16-2928.tar
2014-12-04T19:09:30.000Z	01d5b449c827bf594e5d6bd34a1b9d1b	462	partner_config_download_test.conf
2014-12-04T19:09:30.000Z	eae72e9cb9f93299f9ba749b0d2747a9	462	partner_config_download_test.conf~

Partner Tools



ACADEMIC
PRESERVATION TRUST

Bag Validation

Validate a bag before you upload it.

Get specific details about what to fix.

Partner Tools



ACADEMIC
PRESERVATION TRUST

Bag Validation

```
1. bash
d-128-193-157:bin apd4n$ apt_validate ../testdata/example.edu.sample_no_md5_manifest.tar
[FAIL] ../testdata/example.edu.sample_no_md5_manifest.tar is not valid for the following reasons:
    Required checksum file manifest-md5.txt is missing.

The full APTTrust bagit specification is available at
https://sites.google.com/a/aptrust.org/aptrust-wiki/technical-documentation/processing-ingest/aptrust-bagit-profile
d-128-193-157:bin apd4n$
d-128-193-157:bin apd4n$
d-128-193-157:bin apd4n$ apt_validate ../testdata/example.edu.sample_good
[PASS] ../testdata/example.edu.sample_good is a valid APTTrust bag
d-128-193-157:bin apd4n$
```

Partner Tools



ACADEMIC
PRESERVATION TRUST

Bag Management & Validation

Download from the APTrust wiki

<https://sites.google.com/a/aptrust.org/aptrust-wiki/home/partner-tools>



APTrust



ACADEMIC
PRESERVATION TRUST

Repository

<https://repository.aptrust.org>

Test Site

<http://test.aptrust.org>

Wiki

<https://sites.google.com/a/aptrust.org/aptrust-wiki/>

Help

help@aptrust.org



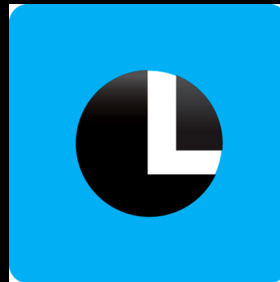
ACADEMIC
PRESERVATION TRUST

Learning to Run

University of Cincinnati Libraries path to
contributing content to APTrust

Digital Collection Environment

- 3 digital repositories
 - DSpace -- UC Digital Resource Commons
 - LUNA -- UC Libraries Image and Media Collections
 - Hydra -- Scholar@UC Institutional Repository
- 1 website
- 1 Storage Area Network file system



Workflow Considerations

- Preservation vs. Recovery
 - platform-specific bags or platform-agnostic bags?
 - derivative files
- Bagging level
 - record level or collection level?

UC Libraries Workflow Choices

- Recovery is as important as preservation
 - Include platform specific files and derivatives
- Track platform provenance
 - use bag naming convention or additional fields in bag-info.text
- Track collection membership
 - use bag naming convention or additional fields in bag-info.text

Bagging (crawling)

1st iteration

- Start with DSpace, Cincinnati Subway and Street Improvements collection
- Replication Task Suite to produce Bags
 - one bag per record
- Bash scripts to transform to APTrust Bags
- Bash scripts to upload to test repository
- Worked well, but...

Scripts available at https://github.com/ntallman/APTrust_tools.

Bagging (walking)

2nd iteration

- Same starting place
- Export collection as Simple Archive Format
 - ~ 8,800 records, ~ 375 GB
- Create bags using bagit-python
 - 250 GB is bag-size upper limit

Bagging (walking)

2nd iteration (continued)

- Collection had three series
 - two series under 250 GB
 - each series was put into a single bag
 - one series over 250 GB
 - used a multi-part bag to send to APTrust, APTrust merges split bags into one
- Bash script to upload bags to production repository

Bagging (running)

3rd iteration

- Testing of tools created by Jamie Little of University of Miami
- Jamie can tell you more but...

Long-Term Strategy (marathons)

- Bag for preservation and recovery
- For most content
 - one bag per collection
 - except for Scholar@UC
 - manual curation and bagging
 - except for Scholar@UC
- Use bag-info.txt to capture platform provenance and collection membership



ACADEMIC
PRESERVATION TRUST

at

UNIVERSITY OF MIAMI
LIBRARIES



UM Digital Collections Environment

- File system based long-term storage
- CONTENTdm for public access



Creating a workflow

- Began at the command line working with LOC's Baglt Java Library
- Bash scripts
- Our initial testing used technical criteria (size, structural complexity, file types)

What to put in the bag?

- CONTENTdm metadata
- FITS technical metadata
- We maintained our file system structure in the bags

Beginning APTrust Testing

- Began uploading whole collections as bags
- Early attempts at a user interface using Zenity

UM's APTrust Automation

- Sinatra based server application
- HTML & JavaScript interface

Starting Screen

UM Libraries APTrust Automation

Let's make some bags.

You've got 3727 gigabytes of free storage.
You probably shouldn't download a collection bigger than 931.75 gigabytes.

Next

Choosing a Repository ID

UM Libraries APTTrust Automation

Please choose which repository you would like to work with.

The repository component indicates which library repository the file is a part of, and serves as a top-level folder in the archive server directory structure.

chc - Cuban Heritage Collection ▼

Repository

arc - Architecture Library

arh

asm - Special Collections

asu - University Archives

bio

chc - Cuban Heritage Collection

cta

dlp - Digital Library Projects

exh

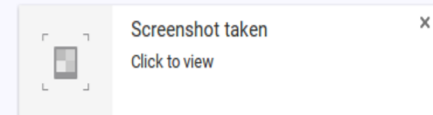
mar - Marine Sciences/RSMAS Library

mus - Music Library

scp - Scholarly Publications

Next

Previous



Choosing Files

UM Libraries APTrust Automation

Click 'View Directories' to select what files you would like to send.

Select All ☐

- ☒ /masters/chc/0003/chc00030000170001001.tif
- ☒ /masters/chc/0003/chc00030000110002001.tif
- ☐ /masters/chc/0003/chc00030000100002001.tif
- ☐ /masters/chc/0003/chc00030000080002001.tif
- ☐ /masters/chc/0003/chc00030000160002001.tif
- ☐ /masters/chc/0003/chc00030000160001001.tif
- ☐ /masters/chc/0003/chc00030000110001001.tif
- ☐ /masters/chc/0003/chc00030000080001001.tif
- ☐ /masters/chc/0003/chc00030000120002001.tif
- ☐ /masters/chc/0003/chc00030000040001001.tif
- ☐ /masters/chc/0003/chc00030000090002001.tif
- ☐ /masters/chc/0003/chc00030000170002001.tif
- ☐ /masters/chc/0003/chc00030000050001001.tif
- ☐ /masters/chc/0003/chc00030000090001001.tif
- ☐ /masters/chc/0003/chc00030000010001001.tif
- ☐ /masters/chc/0003/chc00030000060001001.tif
- ☐ /masters/chc/0003/chc00030000130001001.tif
- ☐ /masters/chc/0003/chc00030000030001001.tif
- ☐ /masters/chc/0003/chc00030000150001001.tif
- ☐ /masters/chc/0003/chc00030000100001001.tif

Viewing Download Results

UM Libraries APTTrust Automation

The collection was downloaded. You can view the details below.

Next

```
receiving incremental file list
```

0	0%	0.00kB/s	0:00:00	(xfr#0, to-chk=3/4)
0	0%	0.00kB/s	0:00:00	(xfr#0, to-chk=2/4)
0	0%	0.00kB/s	0:00:00	(xfr#0, to-chk=1/4)
0	0%	0.00kB/s	0:00:00	(xfr#0, to-chk=0/4)

```
receiving incremental file list
```

0	0%	0.00kB/s	0:00:00	(xfr#0, to-chk=3/4)
0	0%	0.00kB/s	0:00:00	(xfr#0, to-chk=2/4)
0	0%	0.00kB/s	0:00:00	(xfr#0, to-chk=1/4)
0	0%	0.00kB/s	0:00:00	(xfr#0, to-chk=0/4)

Final Upload

UM Libraries APTTrust Automation

Click OK to upload the bag to APTTrust's receiving bucket.

OK

Bagging Strategies

Moved from bag per collection strategy to bag per file

Changed interface to allow selection of individual files

APTrust Benefits for Developers

- Wide range of open source tools available
- Uploading to APTrust is no different than uploading to S3

Where to find us

github.com/UMiamiLibraries

 **LIBRARIES** Digital Collections

<http://merrick.library.miami.edu/>

Why are we doing this again?

- Libraries' and Archives' historic commitment to long-term preservation is the *raison d'être*, a core justification, for our support of repositories. It's why our stakeholders trust us and will place their content in our repositories.
- We expect APTTrust to be our long-term preservation repository to meet our responsibilities to our stakeholders.

Why are we doing this again?

- We know from some experience, and reports from our peers, that backup procedures, while still essential, are fraught with holes.
- We expect APTrust to be our failover if any near-term restore from backup proves incomplete.

UC Libraries, Digital Collections and Repositories

Linda Newman
Head, Digital Collections
and Repositories
linda.newman@uc.edu

Nathan Tallman
Digital Content Strategist
nathan.tallman@uc.edu

<http://digital.libraries.uc.edu>

<https://github.com/uclibs>

<https://scholar.uc.edu>