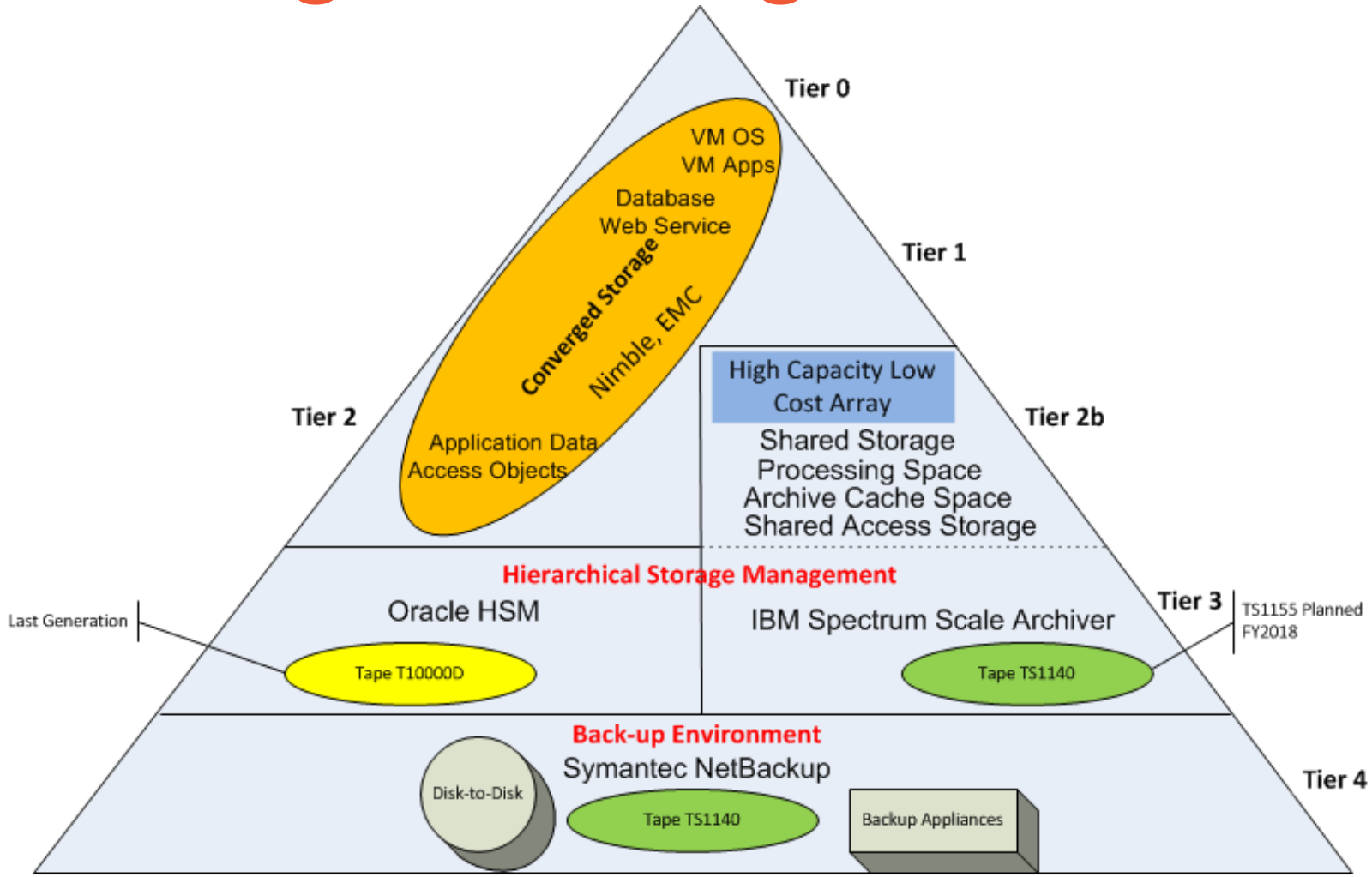
Library of Congress Storage Environment
Update 2018

Carl Watts
Information Technology Specialist
IT Services Operations / Operations and Maintenance / Unix Systems

Converged Storage Tiers

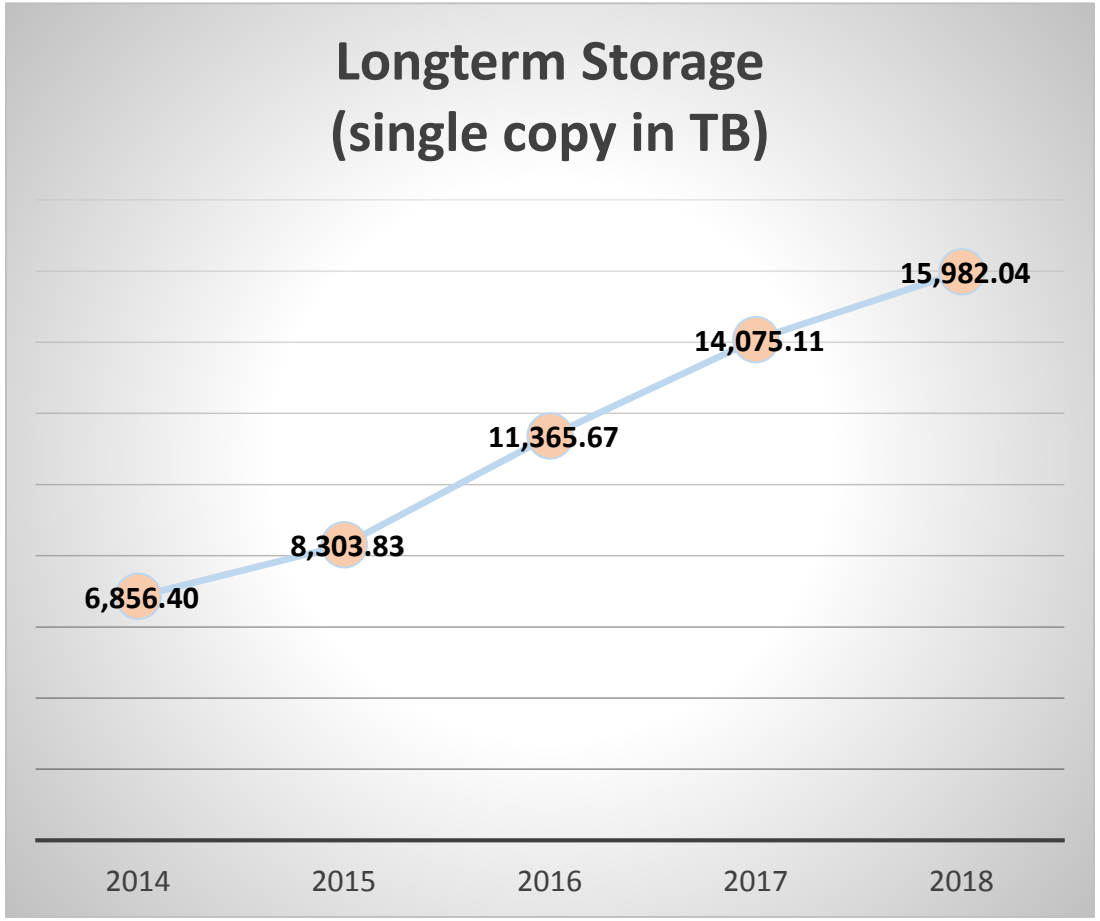


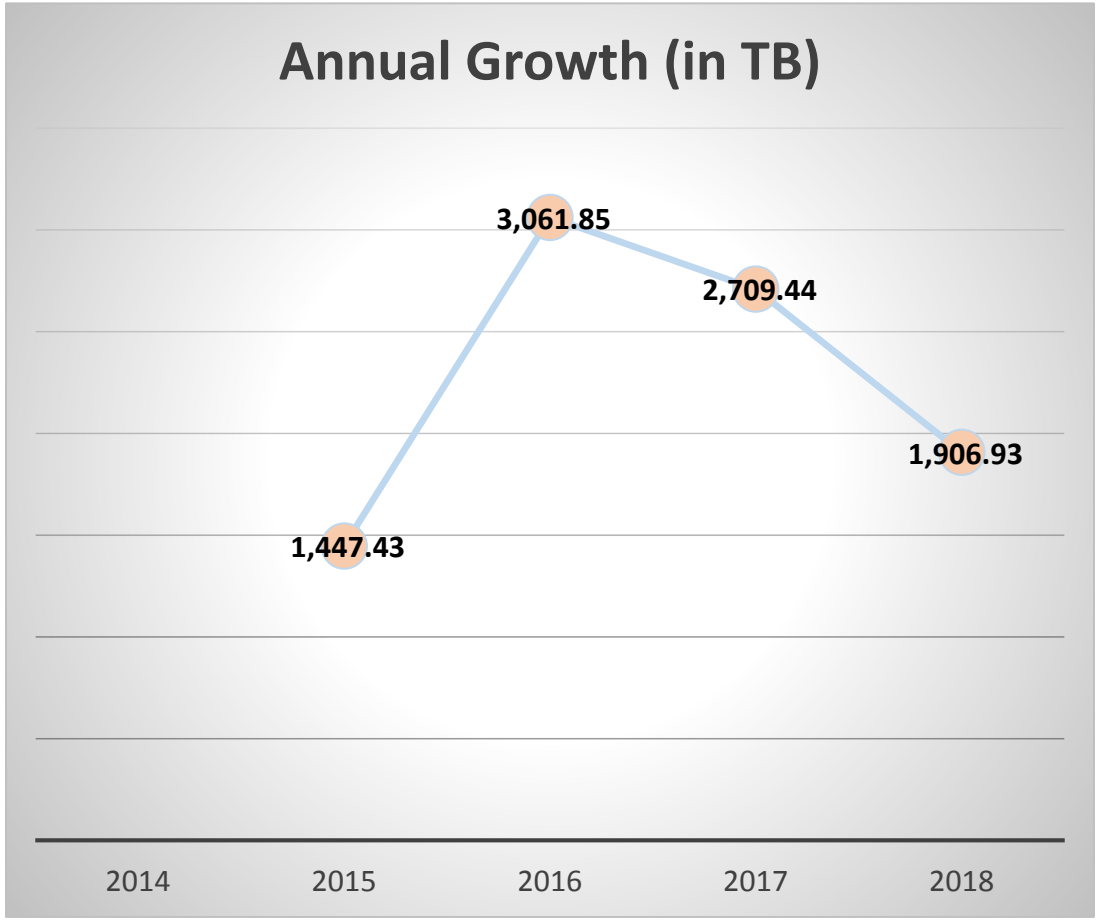
Content Storage

- ***Content*** is equal to single copy of a digital object and it's associated derivative(s)
- Preservation Copies (currently)
  - Standard Collections – two (2) copies distributed across two (2) datacenters
  - Special Collections – two (2) different platforms holding two (2) copies distributed across two (2) datacenters
- Presentation Copies
  - Currently single online copy
  - Near future – two (2) copies across (2) datacenters
  - Future – multiple copies across datacenters and "cloud" providers

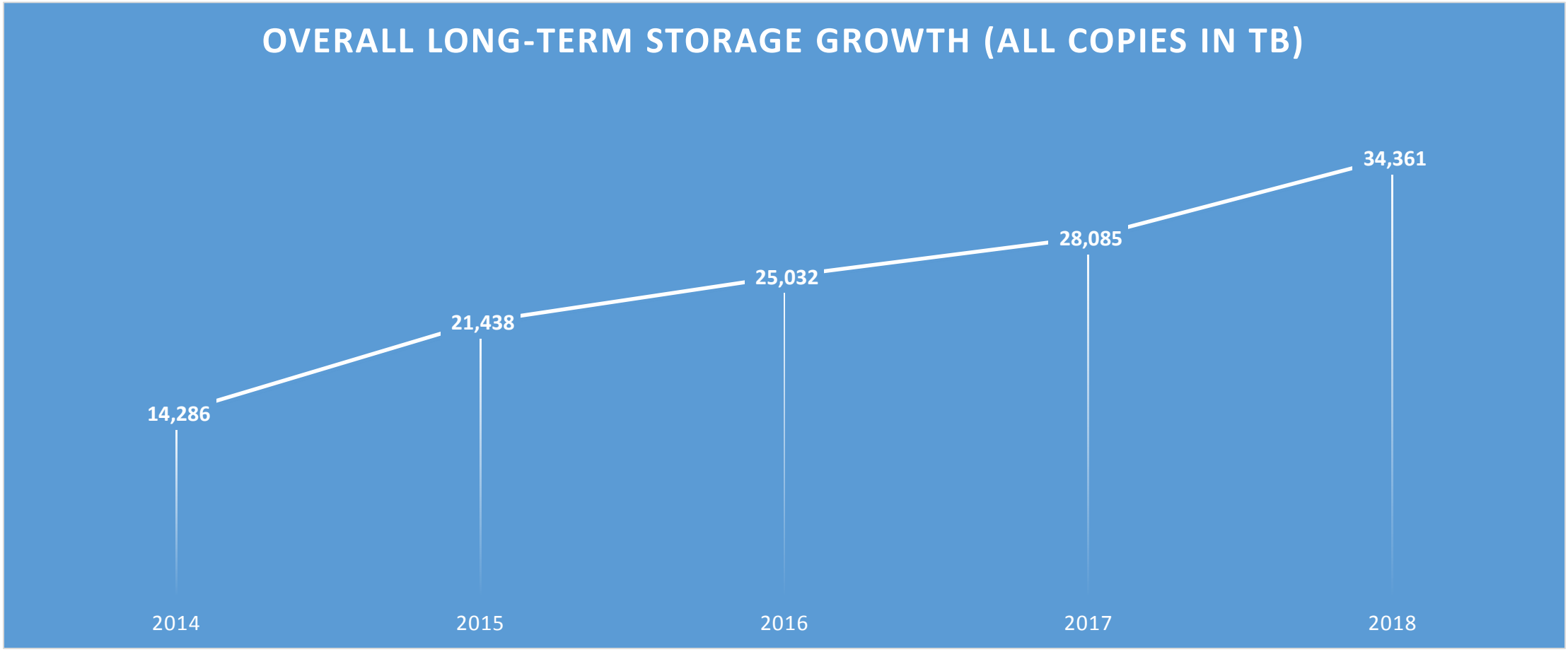
Content Growth – Preservation

Unique File Count:
410M Total Files



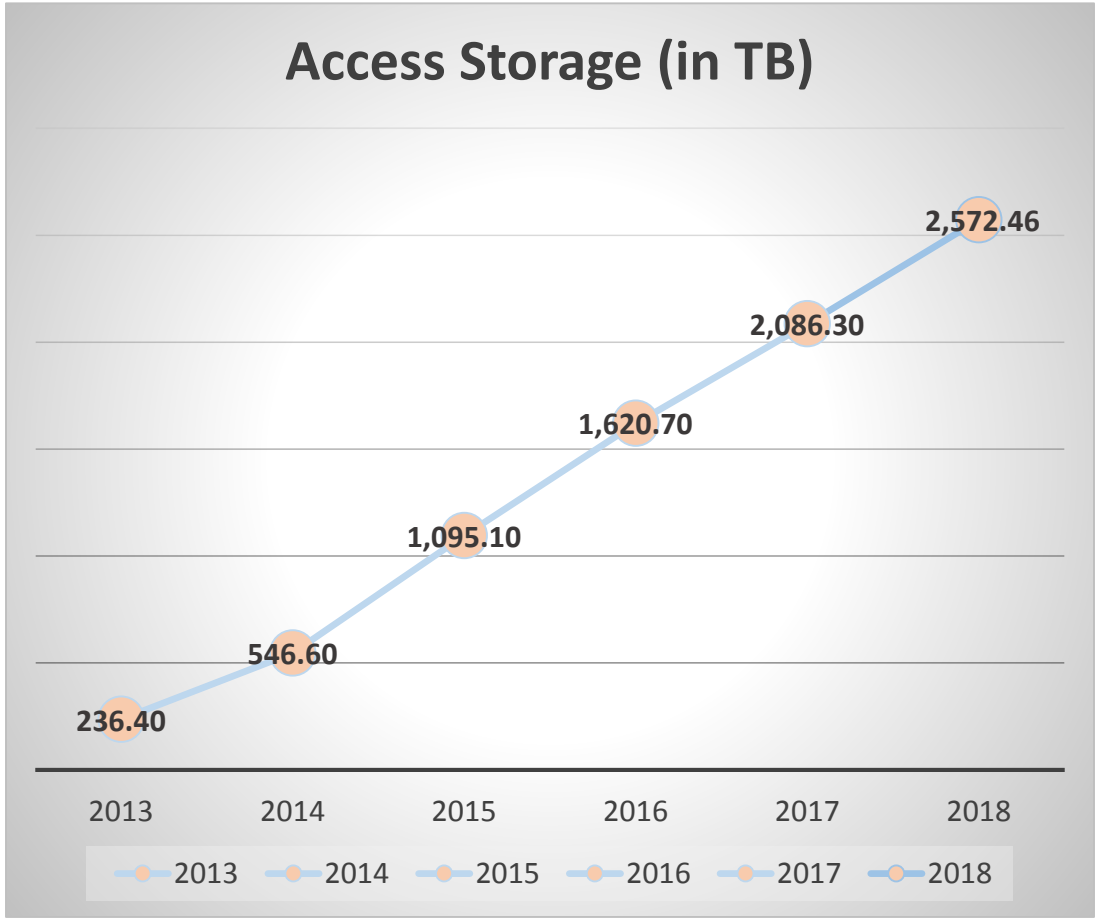


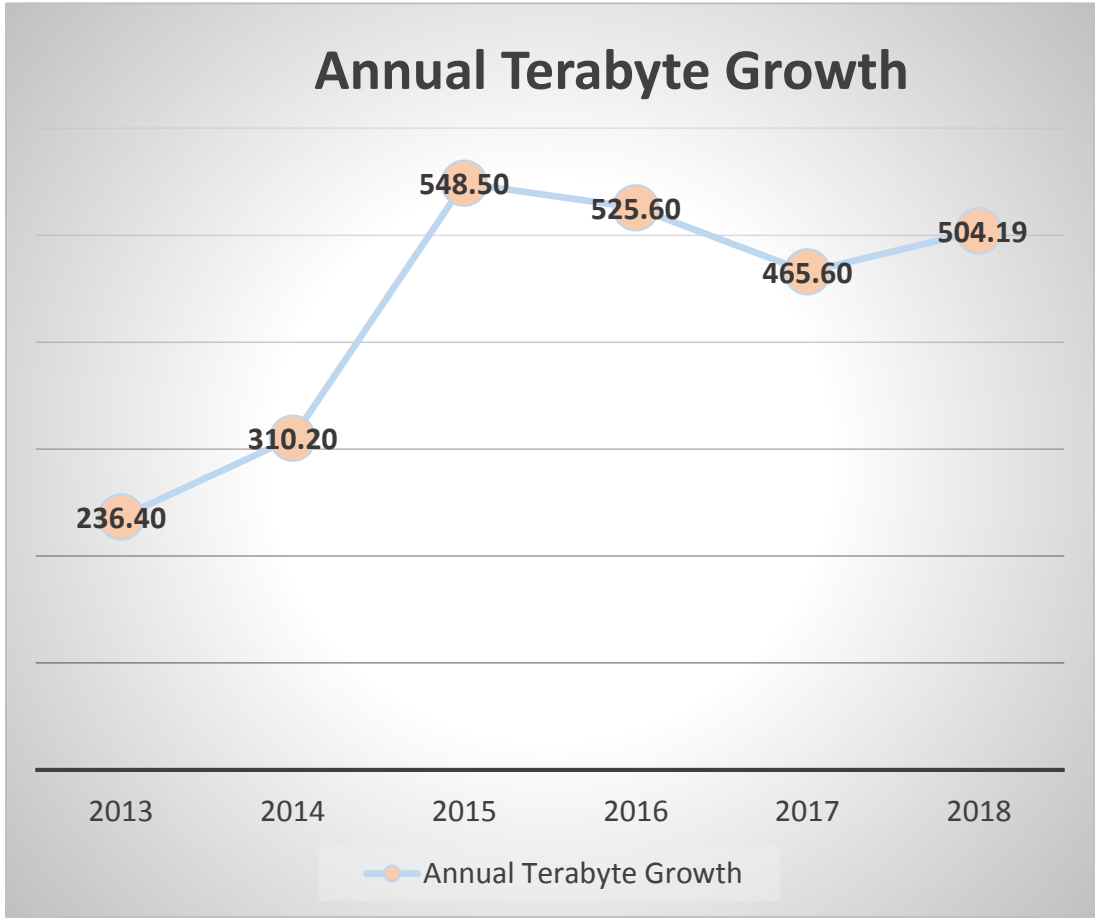
Content Growth – Preservation



Content Growth – Presentation

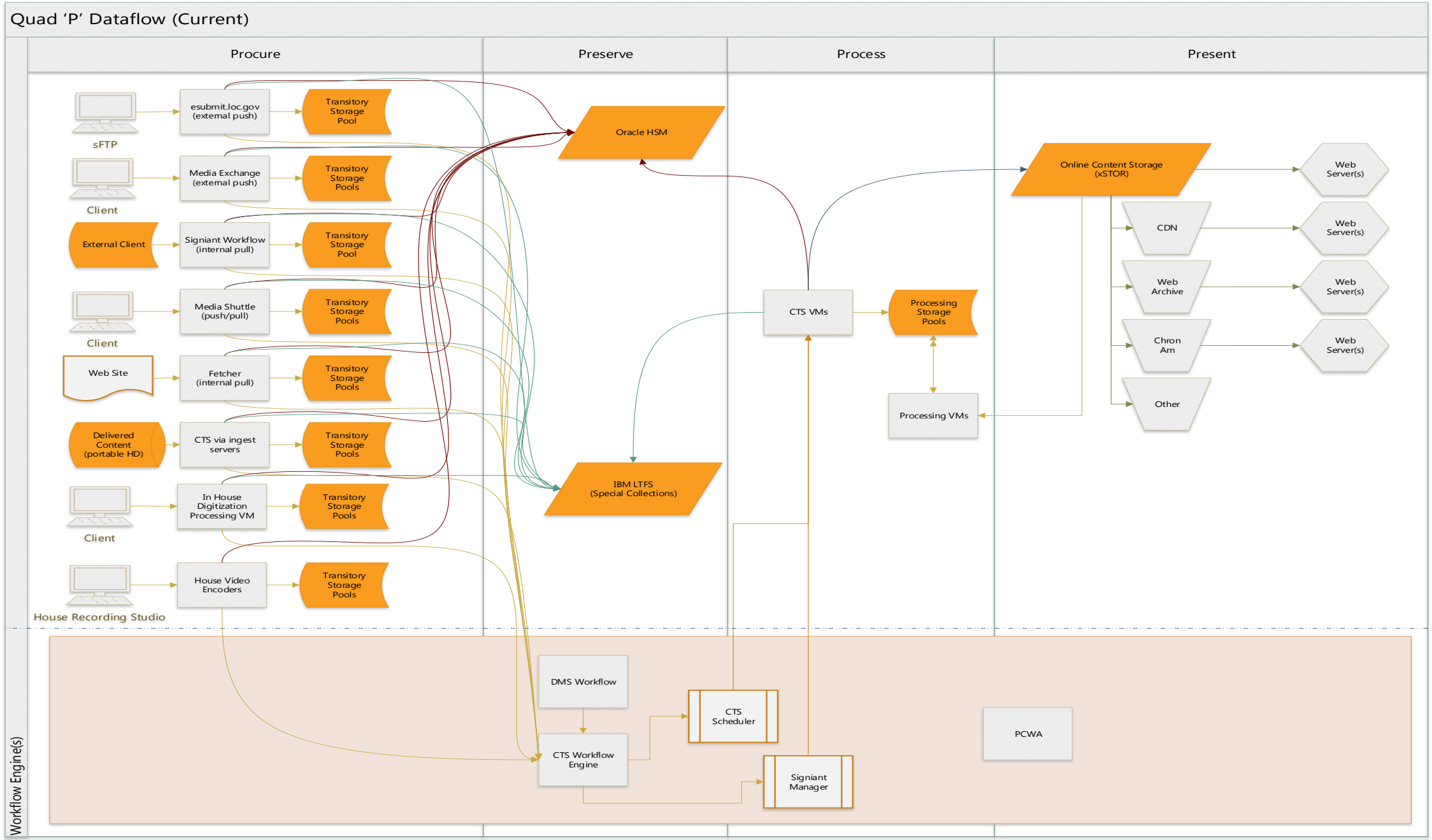
Unique File Count:
344M Total Files





Migrations Continue

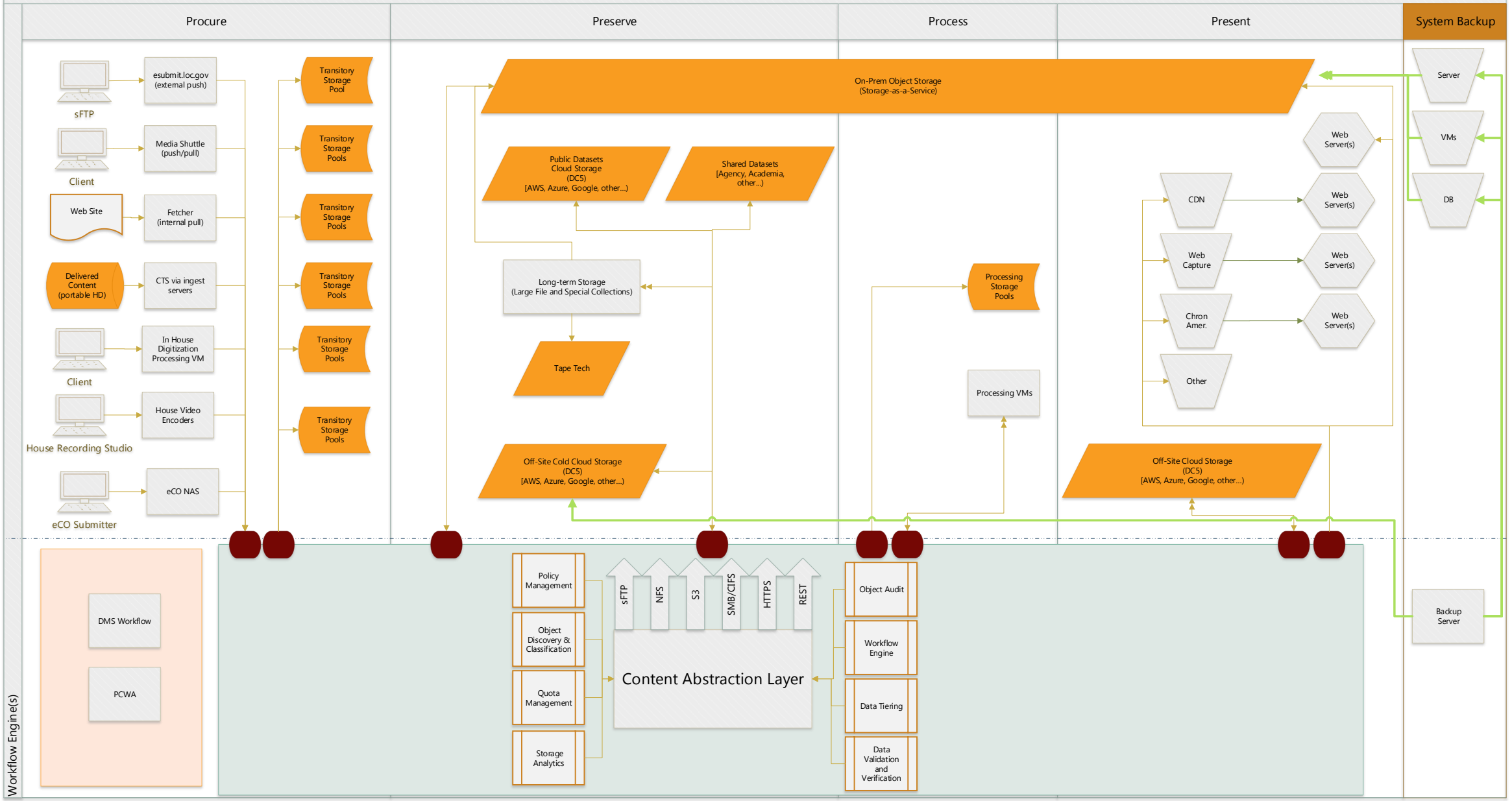
- Consolidating Preservation Storage
  - Combining resource to reduce cost
- Migrating Data Centers
  - Completed migration of presentation storage to new location and system
  - Preparing to replicate data to new data center (2019)
- Migrating Tape Technology
  - Preparing to migrate IBM TS1140 tape to TS1155 tape (2019)

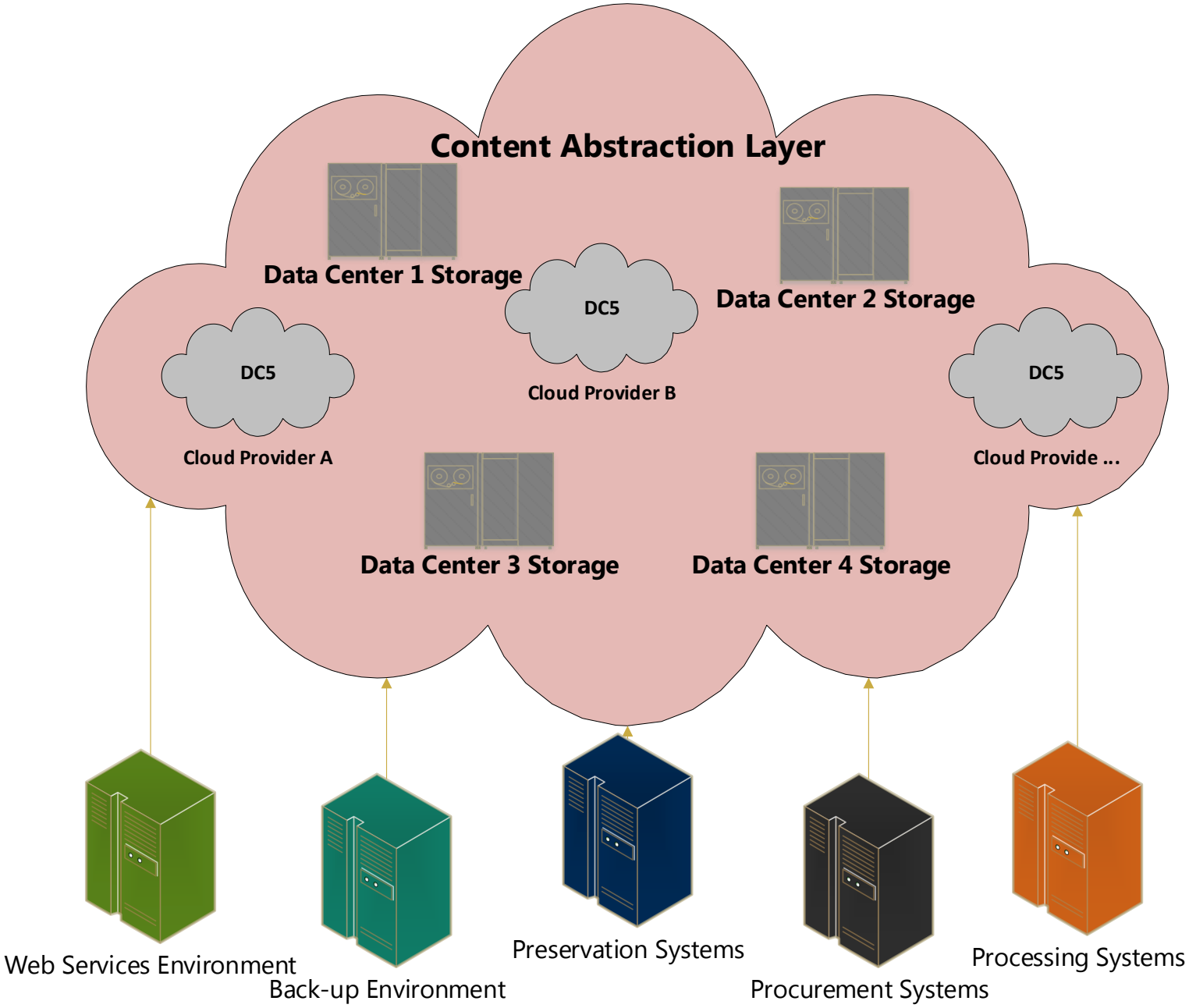

Quad 'P' Dataflow (Current)
Procure
Preserve
Process
Present
esubmit.loc.gov (external push)
Transitory Storage Pool
sFTP
Media Exchange (external push)
Transitory Storage Pools
Client
External Client
Signiant Workflow (internal pull)
Transitory Storage Pool
Media Shuttle (push/pull)
Transitory Storage Pools
Client
Web Site
Fetcher (internal pull)
Transitory Storage Pools
Delivered Content (portable HD)
CTS via ingest servers
Transitory Storage Pools
In House Digitization Processing VM
Transitory Storage Pools
Client
House Video Encoders
Transitory Storage Pools
House Recording Studio
Oracle HSM
IBM LTFS (Special Collections)
CTS VMs
Processing Storage Pools
Processing VMs
Online Content Storage (xSTOR)
CDN
Web Archive
Chron Am
Other
Web Server(s)
Web Server(s)
Web Server(s)
Web Server(s)
Workflow Engine(s)
DMS Workflow
CTS Workflow Engine
CTS Scheduler
Signiant Manager
PCWA

Looking to add Content Abstraction Layer

- Content Abstraction Layer (CAL) would provide:
  - Manage the procurement of data from multiple sources
  - Manage the preservation of content:
    - File fixity checking
    - File validation / usability
  - Manage the automation of content processing
  - Manage the movement / orchestration of data across multiple
    - Systems
    - Data centers
    - Cloud providers
    - External entities
  - Provide a persistent namespace and access method to data

Quad 'P' Dataflow (Proposed)




Content Abstraction Layer
Data Center 1 Storage
DC5
Cloud Provider B
Data Center 2 Storage
DC5
Cloud Provider A
DC5
Cloud Provide ...
Data Center 3 Storage
Data Center 4 Storage
Web Services Environment
Back-up Environment
Preservation Systems
Procurement Systems
Processing Systems

Thank you