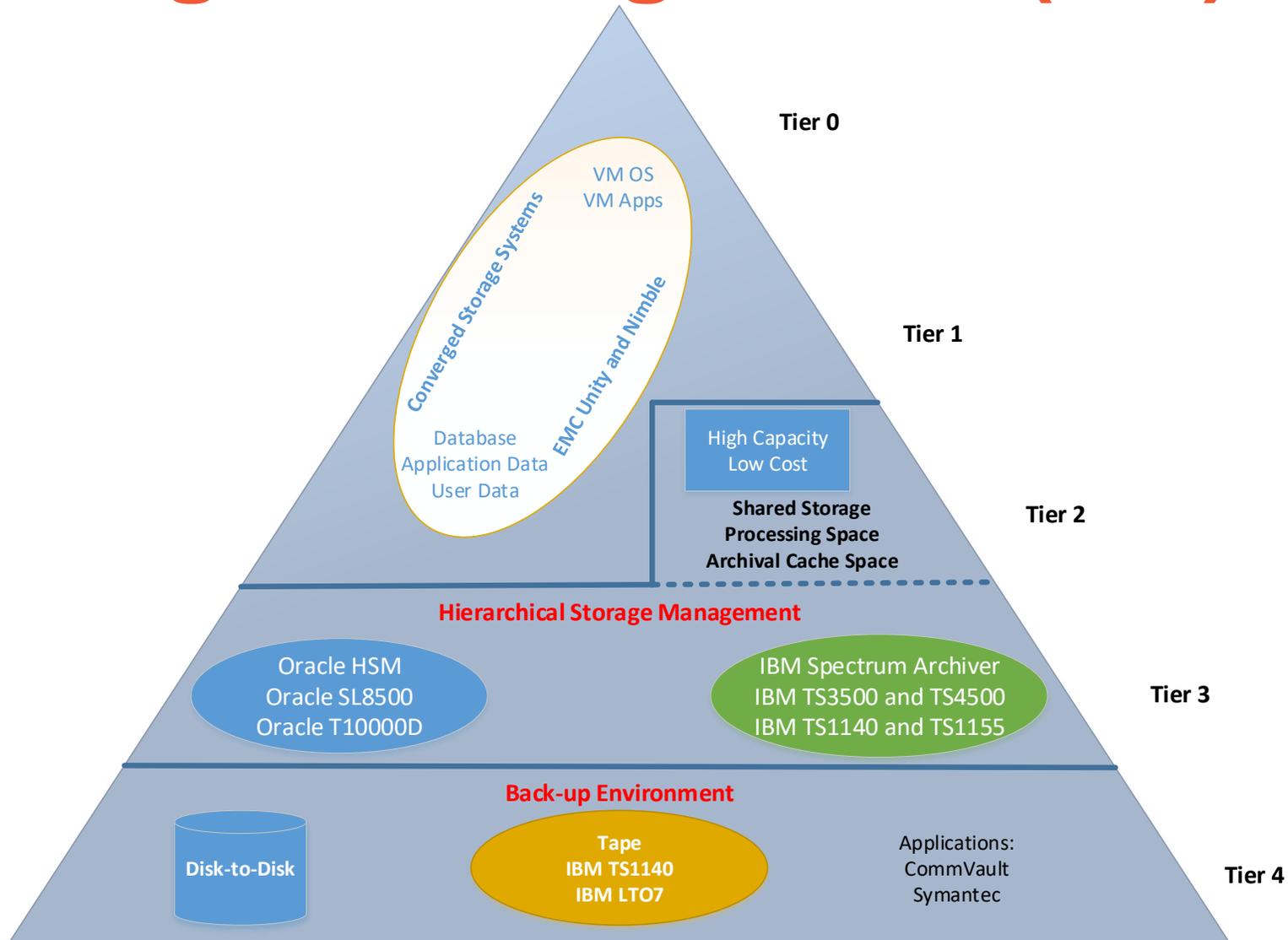# Library of Congress Storage Environment
Update 2019

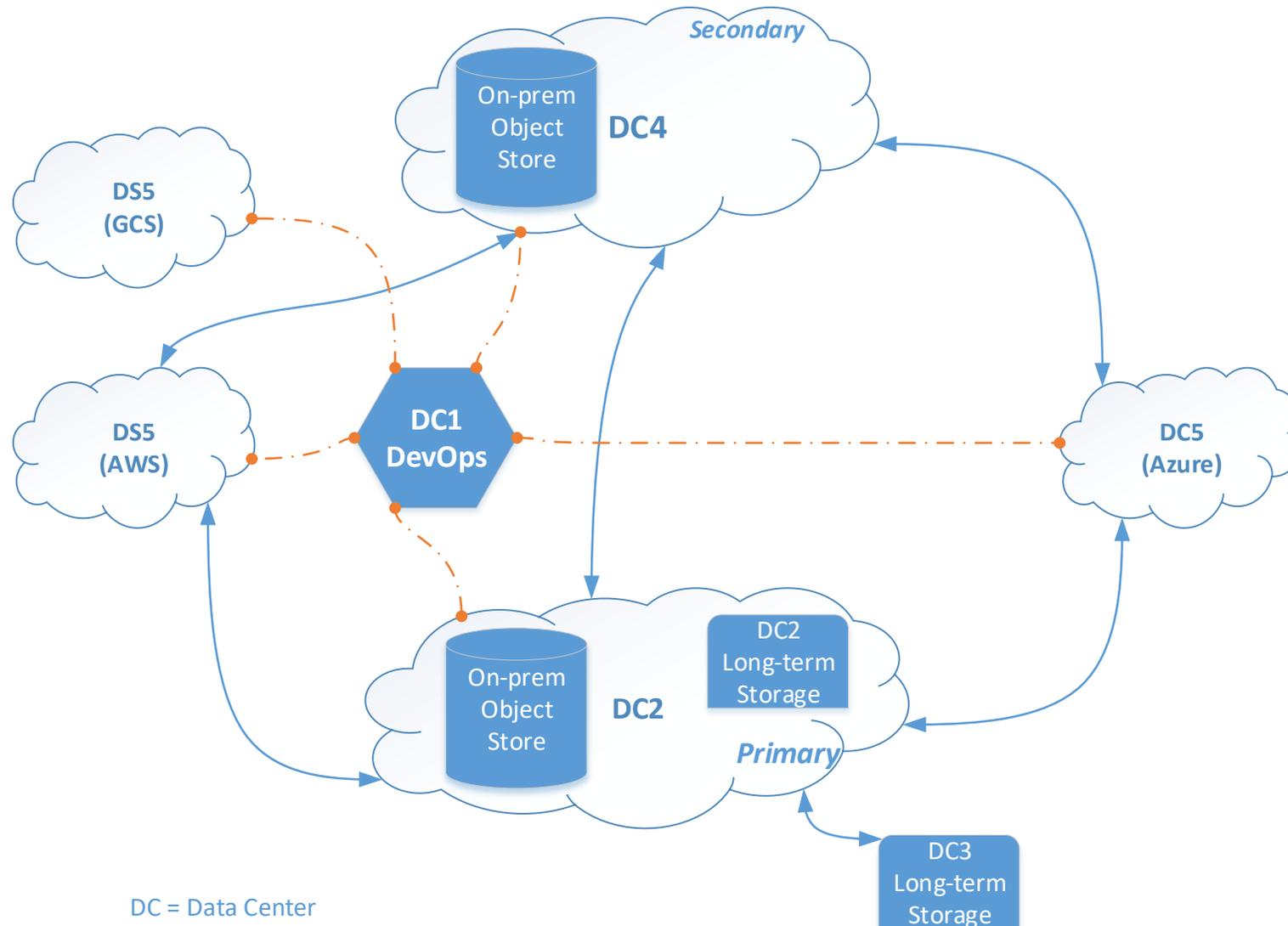Carl Watts
Information Technology Specialist
IT Services Operations / Operations and Maintenance / Unix Systems

Converged Storage Tiers (old)

Tier 0

VM OS
VM Apps

Converged Storage Systems

EMC Unity and Nimble

Database
Application Data
User Data

Tier 1

High Capacity
Low Cost

**Shared Storage
Processing Space
Archival Cache Space**

Tier 2

Hierarchical Storage Management

Oracle HSM
Oracle SL8500
Oracle T10000D

IBM Spectrum Archiver
IBM TS3500 and TS4500
IBM TS1140 and TS1155

Tier 3

Back-up Environment

Disk-to-Disk

**Tape
IBM TS1140
IBM LTO7**

Applications:
CommVault
Symantec

Tier 4

Converged Data Center

DC = Data Center

September 2019

3

Content Growth – Preservation

Unique File Count:
536M Total Files

Longterm Storage (single copy in TiB)

- 2014: 6,856.40
- 2015: 8,273.92
- 2016: 10,955.96
- 2017: 13,956.69
- 2018: 16,448.03
- 2019: 19,504.30

Annual Growth (in TiB)

- 2015: 1,417.52
- 2016: 2,682.04
- 2017: 3,000.73
- 2018: 2,491.34
- 2019: 3,056.27

LIBRARY LIBRARY OF CONGRESS
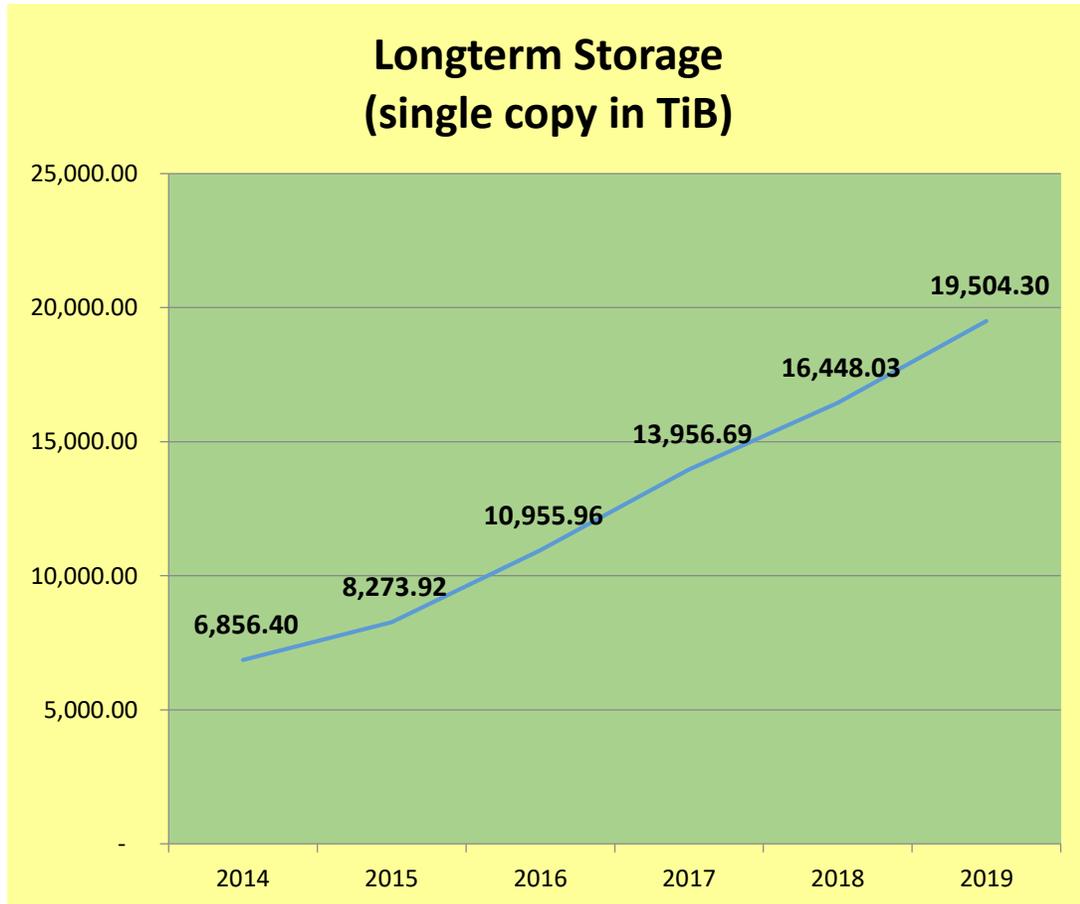
Content Growth – Preservation

About 18% annual growth

OVERALL LONG-TERM STORAGE   (ALL COPIES IN TIB)

- 2014: 14,741
- 2015: 17,789
- 2016: 23,555
- 2017: 30,007
- 2018: 35,363
- 2019: 41,934

LIBRARY OF CONGRESS
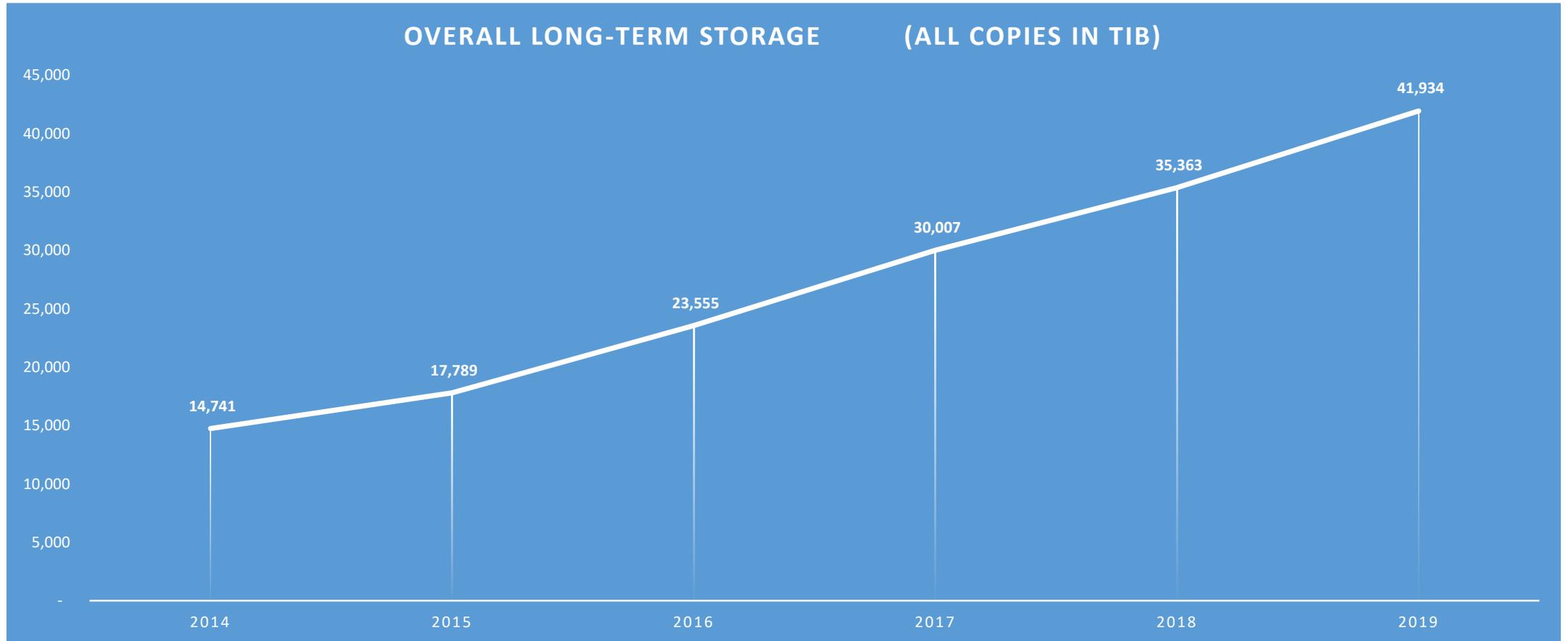
Content Growth – Presentation

Unique File Count: 374M Total Files

About 18% annual growth

Access Storage (in TiB)

- 3,500.00
- 3,000.00
- 2,500.00
- 2,000.00
- 1,500.00
- 1,000.00
- 500.00
- -

236.40
546.60
1,095.10
1,620.70
2,086.30
2,624.99
3,069.88

2013 2014 2015 2016 2017 2018 2019

Annual Terabyte Growth

- 600.00
- 500.00
- 400.00
- 300.00
- 200.00
- 100.00
- -

236.40
310.20
548.50
525.60
465.60
538.69
444.88

2013 2014 2015 2016 2017 2018 2019

September 2019

Migrations Continue

➢ Data Migrations and propagation are now at a constant churn

➢ Completed the Consolidation of Preservation Storage
  - Combined resource of three data center to two to reduce cost

➢ Preparing to Propagate Data Center 2 to Data Center 4
  - Preparing to replicate data to new data center once high-bandwidth network is established

➢ Migrating Tape Technology
  - Migrated one copy of IBM TS1140 tape to TS1155 tape (2019)

➢ Propagating Access Storage to AWS
  - Completed 24 of 48 AWS Snowball transfers (over 1.4PB moved in two months)
  - Project is ongoing and should be completed in six weeks

➢ Propagating Preservation Storage to AWS and Azure
  - Project to start early December to propagate all preservation content to AWS via Snowmobile and Azure via Data Box Heavies

Building out Content Abstraction Layer

➢Installation of StrongLink which will become the Content Abstraction Layer

➢Content Abstraction Layer (CAL) will provide:

- Provide a persistent namespace and access method to data
- Management of the curated data
- Manage the file fixity and fixity checking
- Manage the automation of content processing
- Manage the movement / orchestration of data across multiple
    - Systems
    - Data centers
    - Cloud providers
    - External entities
- Manage the data migration between old and new storage platforms

Adding On-Prem Cloud Type Storage (STaaS)

- Acquiring a Storage-as-a-Service (STaaS)
  - Replaces:
    - active archive Oracle HSM and Spectrum Archive
    - access storage Spectrum Scale
  - Cost is equal to cloud storage vendor cool/cold (about $2/TB per month)
  - Access is about the same current NAS and better then HSM products in use
  - No egress fees for access
  - Managed solutions, which lowers staff administration requirements

Content Storage

➢ ***Content*** is equal to single copy of a digital object and its 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

Quad 'P' Dataflow (Proposed)

September 2018

12

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

LIBRARY LIBRARY OF CONGRESS