

INTERNET2

2022
TECHNOLOGY
exchange

COPING WITH THE DEATH OF UNLIMITED STORAGE

JOHN BAILEY, Washington University in St. Louis

HELEN HOCKX-YU, University of Notre Dame

HELLEN ZZIWA, Harvard University

BOB FLYNN, Internet2 (Moderator)



Agenda

- How We Got Here
- Migration/Reduction Strategies
- Data Lifecycle
- Vendor Education
- The Role of the Community

Your Panel



Hellen Zziwa
**Harvard
University**



Helen Hockx-Yu
**University of
Notre Dame**



John Bailey
**Washington
University
in St. Louis**



Your Moderator



Bob Flynn



Program Manager, Cloud
Infrastructure & Platform
Services



How We Got Here



History of Cloud Storage Quotas/Licenses/Account Limits

Google Drive

- April 2012: 5 GB/user
- May 2013: 30 GB/user
- August 2014: Unlimited
- December 2019: Researching charges for accounts and unlimited storage
- February 2021: End of unlimited storage. Change to tiered pricing model

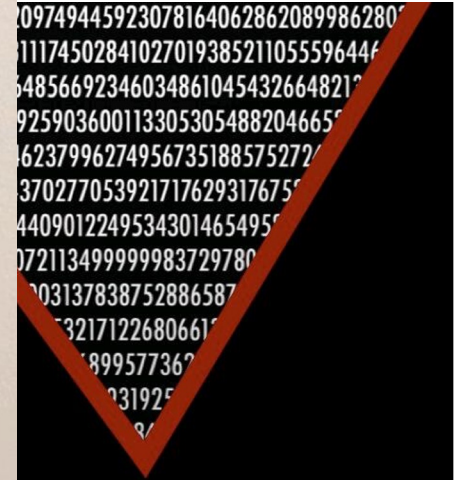
Box

- 2012: 50 GB/user; # users x 2 GB/enterprise
- 2013: 100 GB/user; # users x 4 GB/enterprise
- August 2015: Unlimited
- December 2019: Change to \$820/TB/year pricing model
- Spring 2020: Change to \$130/TB/yea

Microsoft

- September 2013: 7 GB/user
- June 2014: 1TB/user; ??/enterprise
- October 2014: Unlimited
- November 2015: 1TB/user; ??/enterprise
- 2019: Up to 25 TB/user, upon request
- 2019: Many universities move to license certain products for only “knowledge workers”

A Simpler Story



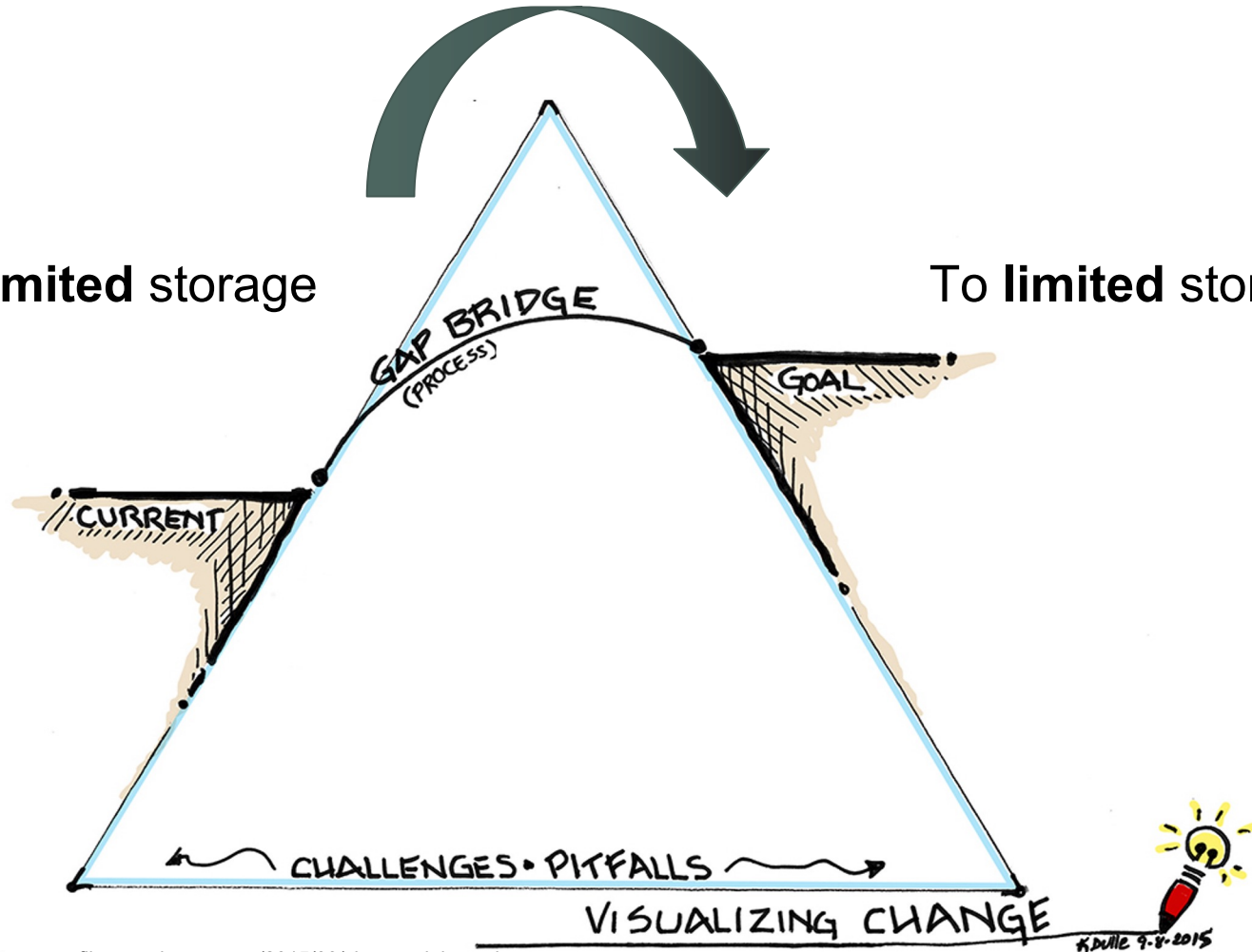
DATA

Migration/Reduction Strategies



From **unlimited** storage

To **limited** storage



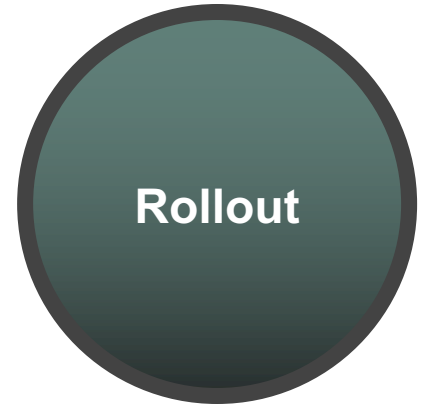
Our approach to bridging the gap



Scope and develop strategies



□ Build and implement solutions



□ Delete or migrate data and transition to operations

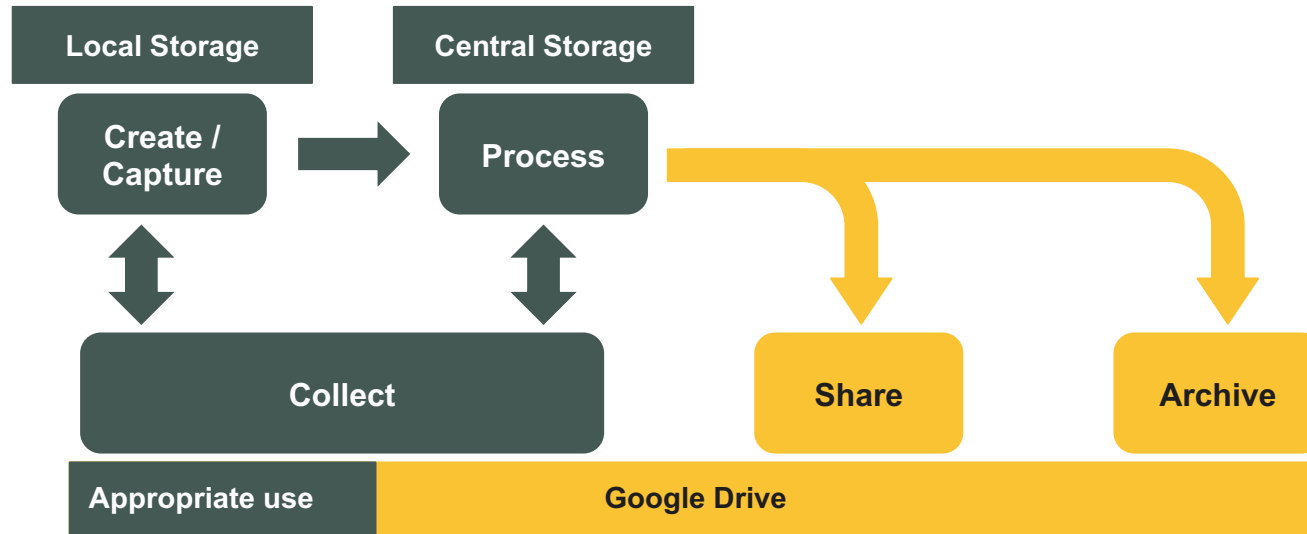
Scope: Storage usage (PB)

0.1

1.9

- **0.4 % users** using **84%** of the storage
- They have use cases that **generate large amounts of data** (Media production, cryo-EM, etc.)
- **Migrate or delete ~1.9 PB** of data in record time OR buy time to figure it out

Storage usage: Example workflow



Destroy?

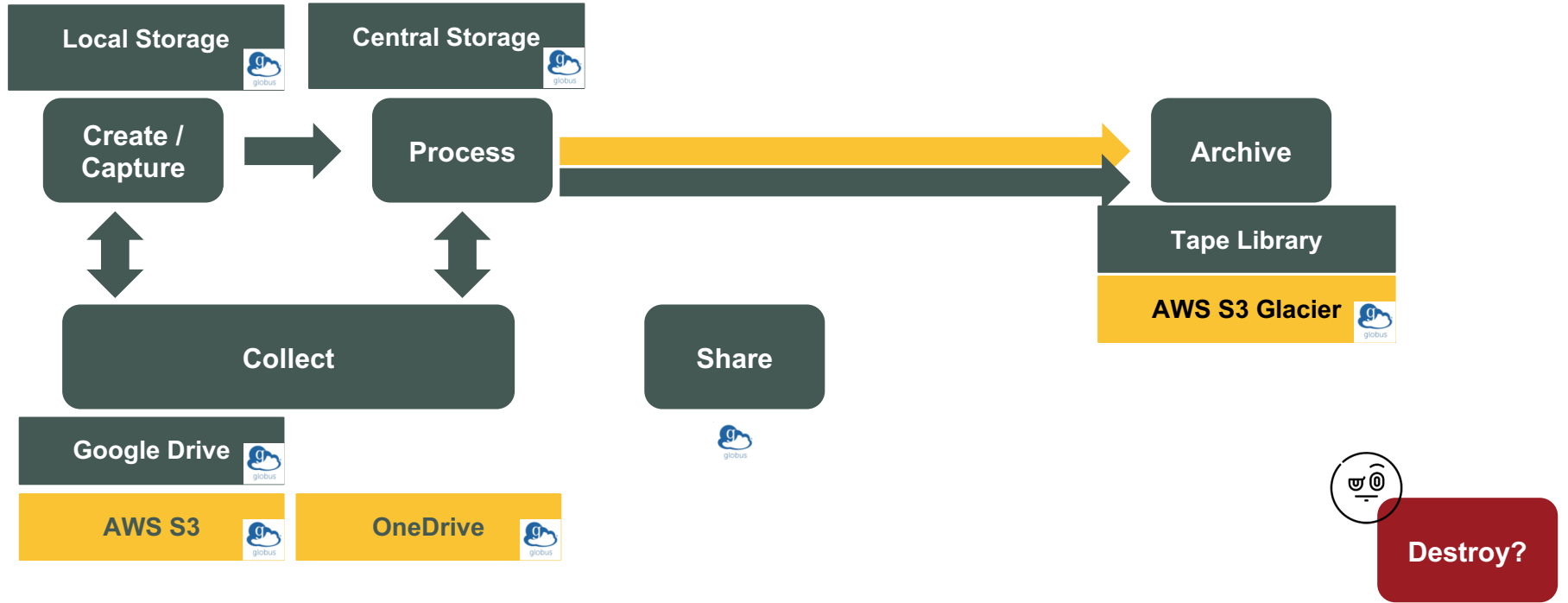
Build: Mitigate storage problem

1.3

0.7

- **Buy time** (until July 2024) **and additional storage** by signing the Internet2 negotiated contract
- Use the time to **prepare to support Google** as an enterprise service

Build: Connect disparate storage locations



Migration of Data Out of Google Workspace (CSWG)

Recommend a NET+ data migration partner spanning multiple platforms

- Define requirements
- Evaluate providers (Globus, CloudM, DryvIQ a NET+ service, Mover.IO, VaultMe, Komprise)
- Recommend option

Explore data analysis tools: (Google Query, Tableau, Sailpoint/Google, Power BI)

Email Dana Voss - dvoss@internet2.edu if you would like to contribute

Storage and Data Lifecycle Management





“Nothing...is ever so expensive as what is offered for free.”

— Viet Thanh Nguyen, *The Sympathizer*



Data Lifecycle

- Birth: data creation, collection, purchase
- Active life: storage, use, sharing
- Inactive: archiving and disposition





Death of Unlimited Storage: Opportunities

- Address overlooked lifecycle stages
- Set strategies and policies
- Drive cultural changes
- Tooling
- Collaboration



Archival Storage Service(s)

“Utilize storage gateways to benefit from Cloud archival storage services for inactive data.”

“Consider moving the on-premise tape library to the Cloud.”

“Work towards a model that will allow more users on campus to benefit from archival storage services.”

“Scope and develop a collaborative archival storage service for research data.”

“Scope and develop a solution for instrumentation storage.”

“Support priorities set by the research governance subcommittee.”

Strategy Area	Strategy Count
Collaborative Environments	1
On-Premise Storage	3
Cloud Infrastructure Storage	3
Directly Attached Storage	1
Archival Storage	3
Research Storage	3
Compliance Storage	4
Storage Expenditure Management	3
Communications and Outreach	3



Disposition of “Orphaned Data”

“Data without a business owner, tied to an individual that has separated from the University”



Guidelines on Disposition of “Orphaned” Data

- Align with existing policies and processes
 - HR normative separation processes requirement: removal of personal content and ownership transfer of business content
- 60 days grace period - existing process allowing retrieval of files
- Additional 365 days retention of shared files
 - Dispose of files not accessed; otherwise offer ownership transfer
- Does not apply to individuals on legal hold
- Does not change existing retention policy



Google Stewardship Campaign

- Overarching message: storage is changing all over the world, not just at Notre Dame
- Build awareness of what it means to be a good ‘Google Steward’
- Conversations with “big” data users
- Timeline January - March 18th
 - Digital Cleanup Day



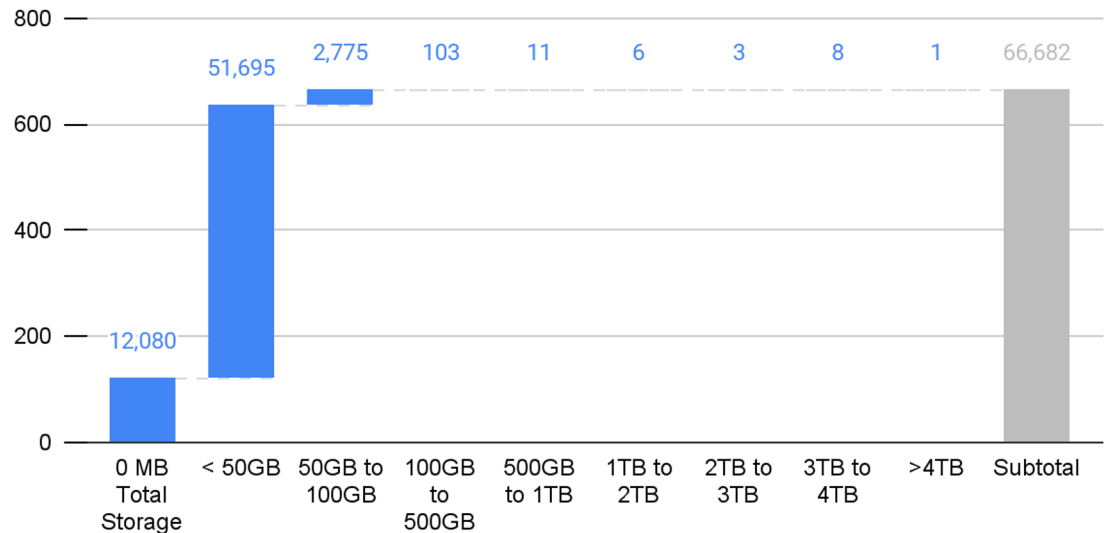


Beyond the Campaign

- Roll out default quota for MyDrive
- Ongoing data curation and archiving

Total Storage Used by and Count of Active MyDrive accounts

Storage amounts reflected in TB





Tooling

- Support all stages of data lifecycle
- Scanning + moving data
- Metadata
- Automation
- Collaboration + handover



Vendor Education



The Near-Death of Unlimited Box Storage at WashU

How WashU Deployed Multiple Strategies Simultaneously

John Bailey

Asst. Director, Cloud Computing
Washington University in St.
Louis

jwbailey@wustl.edu



In The Early Days of Cloud Storage...

- WashU's Box contract provided for unlimited storage.
- Governance committee recommended a reasonable (but still large) cap on storage per user.
- CIO pushed us to embrace a “culture of abundance.”
 - Set each Box user to have no storage quota.
 - Communicated this broadly to the customer community.
 - Actively encouraged customers to move as much data as possible.

Fast Forward 5 years...

- Box informs us that they are going to phase in a new pricing model, charging \$130/TB/year for data over 350TB.
- WashU has 30,000+ active Box users and over 4 PB of data in Box.
- WashU IT begins evaluating options...



Options Considered

1. Push back hard on vendor for better contract terms and/or a built-in archive solution.
2. Evacuate Box and migrate data to SharePoint and OneDrive.
3. Implement a chargeback model to bill departments/schools for the data storage of their heavy users.

What Actually Happened?

- After pushing for better contract terms, we won some concessions (a slower roll-out of the new charges.)
- We developed a charge-back model and began to socialize it with heavy users and their leadership in departments/schools.
- Note: We NEVER stopped pushing Box for a built-in archive solution.
- 1 year later, Box reversed course and re-instated unlimited storage.

Lessons Learned

- Sometimes dragging your feet can turn out to be a good thing!
- Even though you must make plans for what to do if you're cloud storage vendor doesn't budge, never stop pushing on them to do what's right for your institution – including today!

Role of the Community



The Power of Collective Action

- The response to Box: Cancellations, both threatened and fulfilled
- The Internet2 NET+ GWE agreement
- The Cloud Storage Working Group
 - Administrative Tools
- Cloud/Google Workspace Storage Management Working Group
 - Quota Management Working Group
 - Migration of Data Out of Google Workspace
 - Deprovisioning of Accounts
- *What is needed next?*
 - *Lifecycle management best practices?*
 - *More vendor negotiations?*

Discussion

