

Faculty & Staff Advisory IT Committee

Cloud-Based Storage Options

Team

Heather Heckman*, Director, Moving Image Research Collections

Bob Brookshire*, Director of Graduate Studies/Professor, Integrated Information Technology

Juan Caicedo, Undergraduate Student Director/Professor, Civil & Environmental Engineering

Gary Moore, Assistant Dean for Academic Technology, School of Law

Nina Moreno, Associate Professor, Languages, Literatures and Cultures; Linguistics

Ben Torkian, Senior Applications Scientist, Research Computing

Executive Summary

The principal barrier to the adoption of cloud storage services on campus appears to be lack of awareness of existing contracts with cloud vendors. Additional barriers to cloud storage adoption include choice paralysis and legal and privacy restrictions. While some campus users have data storage needs that require enterprise solutions, many more have needs that can be (indeed are) met by solutions at the individual scale. Our aggregated need, however, is great, and satisfaction with OneDrive appears to be low. We recommend that DoIT:

- *Invest in face-to-face communication to faculty and staff about available tools and solutions*
- *Provide regular training and consulting about cloud-storage options available to USC faculty and staff, perhaps in partnership with University Libraries*
- *Encourage use of different services, and continue to expand the menu of cloud storage options available at USC*
- *Work with Counsel to ensure that contractual concerns present no interruption of service for established instances; communicate established failsafes to faculty and staff*
- *Invest in low-cost, scalable, on-premises storage*
- *Purchase a data management system*
- *Provide GSuite and DropBox accounts to campus users*

Enterprise Storage for Research Data

Problem 1: Awareness

Adoption of the cloud services provided through DoIT has been relatively anemic. According to data shared by DoIT, Amazon Web Services (AWS) has seen the most activity, with 8 departmental accounts created. Though there are Azure accounts for each campus in the system, only 2 department-level accounts have been created. No Google Cloud Platform (GCP) accounts have been created.

Our team sent a brief survey about cloud storage for research data to a list of 625 active PIs shared by the Office of the Vice President for Research. It received 94 responses, a response rate of ~15%. More than half of respondents reported storing research data in the cloud, but only 4 reported using AWS, and only 8 reported using GCP. 10 of the respondents with AWS and GCP accounts reported making the accounts themselves; the other 2 didn't know how their accounts were created. This suggests that participation in the campus-wide enterprise-level contracts is *not* the norm.

In written comments, several respondents expressed surprise that cloud storage was available at all for research applications on campus, e.g.:

- As far as I know DoIT does not help create any storage accounts.
- I am not aware of any research data storage options made available by USC. It would be good to have some. I keep all data on lab computers or on cloud accounts paid by grants.
- I am not sure there is such support, this is a discovery for me.

Even among the members of our team, there was confusion about what options were available.

Thus, at this time, the principal barrier to adoption appears to be awareness. Many faculty, staff and students are simply unaware of existing contracts. This is a familiar problem to the Division of Information Technology (DoIT), one that is hardly unique to cloud-storage solutions.

Recommendation 1: Invest in face-to-face communication to faculty and staff about available tools and solutions.

The problem of timely communication in an organization as large and diverse as USC is one that will never be fully resolved.

However, coordination with the Office of the Vice President for Research could result in targeted approaches to users likely to benefit from existing enterprise cloud contracts. More ambitiously, DoIT could aim for annual, in-person appearances at academic unit (or departmental) faculty meetings, where information could be shared about software and services available to faculty more generally.

Problem 2: Choice paralysis

Faculty interested in cloud services can be paralyzed by choice and uncertainty about pricing. There is a fear, not unfounded, of path dependency--that the selection of one cloud provider will make it difficult to change providers in future. An Internet2 comparison of AWS, Azure, and Google Cloud Platform (GCP) highlighted many of these concerns as recurrent themes. Although AWS was described generally as, "Established and familiar; the option with the

greatest brand awareness,” and “Feature rich and reliable,” contributors noted the following weaknesses and questions:¹

- The actual cost is often not articulated clearly and billing can be hard to understand
- There is no predetermined total cost of ownership
- Tools keep being developed and changed at a rate that is hard to keep up with for support purposes and is confusing and intimidating for users
- Why do people find them so trustworthy?
- What do they do better than other cloud providers? Why use AWS vs. others?
- How much more security effort is it to secure something now vs. after the campus has become more experienced?
- How do costs compare, especially for specific features?
- How do we get to pricing predictability?

Relatively fewer strengths, weaknesses, and questions were available across almost all categories considered in the same Internet2 document for Azure and GCP, suggesting that these services are associated with even greater levels of uncertainty. About Azure, contributors noted, “Rollout of changes is often a surprise and there is little warning; the roadmap for the future is often unclear.”² Similarly, of GCP, they wrote, “Roadmap for the future is typically opaque.”

Recommendation 2: Provide regular training and consulting about cloud-storage options available to USC faculty and staff, in partnership with University Libraries.

Although some survey respondents had a clear sense of what they needed and strongly preferred to retain control in the decision-making process (e.g., “I do not want UTS telling me what to do with my data”), others seemed interested in greater guidance:

- How do we store data in the cloud?
- I'm not sure my current methods for data storage are compliant for federal agencies.
- Would like a centralized dedicated team that can enable rapid response for the purposes of grant submission, as well as state of the art management of data and systems when awarded funding.

A consulting service could meet this need without imposing a “one-size-fits-all” solution. At least 2 respondents, including one currently embedded at NSF, requested specifically that such a service be provided by the Libraries. We also note that USC researchers who currently make use of cloud storage for their data could be invited to present at trainings, providing valuable experience and concrete use cases.

¹ <https://spaces.at.internet2.edu/display/CA/Topic+-+Cloud+Provider+Feature+Matrix>, last updated May 2018

² Another Azure weakness noted in the Internet2 report worth highlighting for our campus is surprisingly poor integration with Office365.

Problem 3: Legal barriers & other restrictions

There are significant legal and bureaucratic barriers to adoption, as well. A first level of conflict is posed at the institutional level. The GCP contract, for example, aroused concerns with University Counsel that have delayed the creation of accounts by a calendar year. An additional layer is imposed by third-party retailer vending arrangements, which have caused delays here at USC, and, according to the Internet2 document, at other higher ed institutions. As one survey respondent signaled, such delays can be especially problematic for researchers working on grant deadlines:

Particular fields and research agendas carry further restrictions. Here HIPAA compliance is a highly visible, but not unique, case.

Recommendation 3a: Encourage use of different services, and continue to expand the menu of cloud storage options available at USC.

While any particular department might reasonably wish to work with a single provider for efficiency, we do not want to become dependent on a single provider as a campus. Agreements with multiple vendors are more likely to serve the broad and diverse needs of all our institution's users than any single contract. Central IT is uniquely positioned to cultivate strategic diversification at the institutional level, both through the consulting service recommended above, and through continued exploration of additional vendors. Services already in use by campus researchers include Google Drive, Dropbox, BackBlaze, box, Cloud49, Linode, and Egnyte. Also there are several vendors, such as Lenovo, who recently won a state contract, that provide storage solutions and have great interest in getting a foothold in academia.

Recommendation 3b: Work with Counsel to ensure that contractual concerns present no interruption of service for established instances; communicate established failsafes to faculty and staff.

Diversification can help alleviate these problems at the moment of selection, but there is a further risk at each contract renewal. It is a grave problem if campus researchers do not have a place to put newly generated data; it is arguably a graver one if campus researchers lose access to data they are storing. NSF, for example, requires datasets to be publicly available for at least three years after the end of the grant project; DOIs must not be broken. A working group including representation from faculty, staff, and DoIT might be tasked with this project.

Recommendation 3c: Invest in low-cost, scalable, on-premises storage.

At this time, the rates for the storage tiers offered locally are orders of magnitude higher than the rates charged by cloud providers. Affordable local storage could bridge the gap while researchers wait for approvals and configuration. It would solve other problems, too. Some sensitive and restricted data necessitates local storage; deep storage should be available to

researchers working with such data. Finally, on-premises storage could also allay concerns about egress and path-dependency that remain after adoption.

Recommendation 3d: Purchase a data management system.

Users that have adopted cloud storage at the enterprise level note that there is a critical, and unmet, need for a centralized data management system at USC. Cloud services provide a place to put data, but they do not curate that data. Moreover, researchers might rationally opt to distribute data across multiple platforms (e.g., because of differing levels of restriction, or to mitigate the risk associated with dependence on a single private-sector vendor), greatly increasing the appeal of a single data management system. Even if individual researchers do not wish to distribute data across multiple platforms, we are stronger as a campus if we are able to work with multiple cloud providers. We recommend that the University acquire a data management system that would be administered in cooperation with, or perhaps under the umbrella of, University Libraries.

“Personal” Storage

Many of the researchers surveyed worked with relatively small data sets. The most widely adopted cloud storage solutions were DropBox (n=32) and OneDrive (n=21); regrettably, we did not ask about Google Drive. Moreover, storage of documents and other materials made in the course of standard job duties for faculty and staff poses another area of need for cloud-based storage. Although any individual’s storage footprint might be relatively small (small enough, indeed, to be met by free or very low cost contracts with commercial vendors), our aggregated need across USC is great.

Nina Moreno surveyed her colleagues in Languages, Literatures & Cultures (LLC) and Linguistics areas that unite research programs in the Humanities and the Social Sciences. None of the 26 respondents were using cloud-storage at scale, but the overwhelming majority (25 out of 26) were nonetheless using cloud storage for critical work duties. USC’s faculty and staff use personal cloud accounts for backup, for portability across multiple spaces and devices, and for collaboration. The ease and reliability of freely available accounts from Dropbox and Google often outweigh security concerns. 23 respondents used Dropbox accounts, 20 used Google accounts, and 8 used iCloud accounts. 17 used OneDrive, but simple math suggests that no more than 3 of these could be using OneDrive in isolation.

Problem 4a: OneDrive is an unreliable backup solution

Many respondents to both surveys took the opportunity to voice frustration with OneDrive. Sample comments from the PIs include:

- We are using university Office 365, which is not well suited for research data and external dropbox and box
- The cloud service provided by the university is a waste of money

Of the 17 respondents in LLC who reported they were using OneDrive (often in parallel with another cloud service), 7 shared negative comments, including reports of syncing failures, glitches, and “poor reliability.” Without any snapshotting, for example, OneDrive can be an unforgiving solution.

Problem 4b: Collaboration is difficult with OneDrive and Office365

Another recurring theme related to collaboration. Some PIs noted that inter-institutional collaboration was a strong *disincentive* to use OneDrive storage for their research data:

- I can't share onedrive with co-PIs outside of USC
- OneDrive is a good addition to our capabilities esp in light of it being HIPPA compliant. However, functionality with colleagues off campus is limited.

Faculty in LLC similarly felt that the collaborative features in Office365 were inferior to those available free of charge to Google account holders. Anecdotal reports to our team suggested that while Microsoft Teams meets some of the collaborative needs unmet by OneDrive, it is an incomplete solution. Notably, we authored this report in Google Drive.

Recommendation 4: Provide GSuite and DropBox accounts to campus users.

In recommendation 3a above, we urged DoIT to continue to expand the menu of cloud storage options available to faculty and staff on campus. We believe GSuite and Dropbox, both of which offer solutions targeted at Higher Ed, constitute special cases. The reality is that these are the services our faculty and staff *already use* to backup (or even primarily store) their files, and to work collaboratively.

One PI wrote, “I would love it if USC got an institutional license with some of the cloud storage services I am already using: Evernote, Google One, Dropbox. Alternatives like box.com would be useful too (though I'm unlikely to switch from Evernote to Microsoft OneNote or another Office365 solution).” Similarly, a faculty member in Linguistics and LLC wished that USC would, “start using Google like many other higher ed institutions.” We strongly echo these sentiments.