



**FAS Research Computing
Enterprise Architecture
HUIT Cloud Consulting**

Advisory

Improved Support for Research in the Cloud

THIS DOCUMENT IS IN THE REQUEST FOR COMMENTS PROCESS
FROM 02-26-2021 TO 03-26-2021
RETURN COMMENTS TO ben_rota@harvard.edu

Authors: <ul style="list-style-type: none">• Rota, Ben	Audience Level: <ul style="list-style-type: none">• IT Director / Manager• Solution Architect and Project Manager• Application Developer and Designer• DevOps Staff
Version: 1.02 Last Revised: 6/7/2021 5:19:00 PM (auto updated) Status: RFC closed EOD on 03/26 Document Type: Single Topic Advisory	Distribution Scope: <ul style="list-style-type: none">• Harvard-wide
Workgroup Members: <ul style="list-style-type: none">• Burson, Jefferson• Sevier, Raoul• Yockel, Scott	Reviewers: <ul style="list-style-type: none">• Snyder, Jason• Research Computing & Data Council Executive Committee• HUIT Enterprise Architecture• HUIT Cloud Consulting• Interview participants• Research Computing & Data Council members

Document Change Log

Date	Editor	Changes
02-26-2021	Ben Rota	Initial Draft
03-03-2021	Ben Rota	EA-driven formatting changes
03-08-2021	Ben Rota	Updated Executive Summary, expansion of Other Considerations, extension of RFC period
03-19-2021	Ben Rota	Incorporated feedback from Jason Snyder and Scott Yockel
03-24-2021	Ben Rota	Incorporated feedback from RCDC EC
03-30-2021	Ben Rota	Incorporated feedback from interviewees
03-31-2021	Ben Rota	Grammar fixes, added URL to Appendix A, moved to v1.0
04-14-2021	Bill Barnett	Contextual alignment with Schools
5-27-2021	Ben Rota	Added additional RCDC feedback

1. Executive Summary

Harvard has a large, cost-effective set of on-premises services for researchers. Nevertheless, the University spent well over \$1m for research in the Cloud in CY2020 not inclusive of vendor credits which themselves may represent over \$1m in research spend. The level of support that researchers experience in their use of the Cloud can vary widely. We have identified the following top challenges and recommendations:

- Researchers aren't necessarily aware of the Cloud services that are available to them. We recommend **expansion of existing information resources** (<https://researchsupport.harvard.edu/>, ServiceNow Knowledge articles) to enumerate the services available to researchers and help guide them in their use of the Cloud at their respective schools.
- While it is easy enough for researchers to create a Cloud account on their own, doing so in a compliant and secure way requires significant effort due to the "shared responsibility" security models Cloud providers use. We recommend Harvard create a process to **deliver Cloud accounts and supported Cloud tools** to researchers, delivered through school research computing groups that have as much **security and compliance "baked in"** as possible.
- Truly leveraging the potential of the Cloud for research requires a certain amount of expertise in both Research Computing and in Cloud technology. While Harvard has many staff members with experience in one of the two areas, there are very few who are experts in both, and when they exist, such people aren't necessarily available to all. Local research groups should **expand their portfolio to include Cloud support** and HUIT should expand services to support those teams.

2. Problem Statement

Supporting research in the Cloud effectively requires both expertise in the Cloud and in Research Computing and Data (hereafter referred to as simply "Research Computing"). When researchers decide to use the Cloud, they find themselves needing to largely build up that Cloud expertise, which can include:

- Choosing the right Cloud provider
- Optimally architecting solutions
- Deploying Cloud resources
- Basic systems administration

The result is that they need to spend time or money that could otherwise be used for research getting themselves or their teams up-to-speed on the Cloud. It also means that they're on their own for important issues like security and cost management. Nevertheless, Cloud computing is a growing portion of the University's research spend.

Existing services (notably Consolidated Billing and Enterprise Discounts) are viewed positively, but knowledge of those existing services is not sufficiently widely distributed.

We have identified three key challenges/opportunities in the support of research in the Cloud at Harvard:

- There is no easy way for researchers to understand what is available to them at Harvard regarding research in the Cloud
- When researchers do use the Cloud, they need access to managed platforms and tools that can reduce friction in deploying Cloud tools and thereby reduce their "time to discovery"

Research in the Cloud

- Not all Harvard researchers have access to an organization that understands their research and can provide expertise in both Cloud and Research Computing that they can reach out to for guidance, consulting, or hands-on assistance

3. Requirements and Recommendations

Guiding Principles:

Any solution put into place must be built with the following principles in mind:

- The community that supports researchers at Harvard is broad and varied, and any solution must involve strong collaboration and communication across that community
- “Time to discovery” is of primary importance for researchers in their use of the Cloud
- Research Computing expertise to design and support Cloud solutions usually resides in groups local to the researcher
- The Cloud will not be cost effective if compute and storage assets are provisioned without attention being paid to how Cloud cost management works
- Provisioning assets in a secure and compliant manner must be the path of least resistance for researchers

Key Recommendations:

These recommendations are further detailed in the Discussion section.

- Existing centralized information resources (notably <https://researchsupport.harvard.edu>) should be expanded to include information on Cloud-related services available to researchers from central and local groups
- HUIT, in collaboration with local Research Computing units, should institute generally accessible Cloud account and tool provisioning that provides researchers with environments pre-configured for security and compliance
- Local Research Computing groups should expand their portfolios and skills to include Cloud support. The staff chosen for those roles should work in close coordination with HUIT to deliver consistent Cloud support to the University.

4. Discussion

The methodology used to reach these conclusions included input from the Higher Ed community (via survey – see Appendix A) and from in-person interviews with Harvard research teams (e.g., HMS DBMI), IT staff with explicit research support roles (e.g., FASRC), and more administrative support teams who field researcher requests (e.g., SEAS IT).

For each of the key recommendations, we will discuss methods and requirements for their implementation. We will also attempt to call out where financial investment may be required.

4.1. Centralized location for information

Documentation of existing services/vendors available to researchers should be published in central, easily accessible locations and referenced locally for ease of access.

4.1.1. Required changes

We generally expect that this implementation can be achieved using current staffing levels. Most of the content for back end support will need to be created by HUIT, but collaboration with local Research Computing teams will be required to frame out researcher-facing service offerings.

Research in the Cloud

4.1.2. Implementation approach

We recommend that we provide researchers information in two ways:

- A section on Cloud on <https://researchsupport.harvard.edu>
- Service Now knowledge articles in at least the HUIT-supported instance and likely in others as well, both for discoverability by researchers and in support of local service desks
- A section on Research on the HUIT web page(s)
- Cloud-based Services portfolios on School research computing web sites.

These sources of information should cross-link so that only one location need be updated as services and offerings change.

The information at least needs to include:

- List of IT contacts by school who can help researchers use the Cloud, as well as the value proposition for reaching out to them
- Key research use cases for the Cloud and examples of successful Cloud usage
- Listing of Cloud providers and what services Harvard provides for each, including availability of programs like STRIDES
- Guidance on data sharing in the Cloud
- Resources that guide researchers on whether to use Cloud vs. on-prem vs. some hybrid solution, and further guidance on how to choose a Cloud provider
- Guidance on vendor credits: how to apply for them and how to use them most effectively
- Guidance on cost management in the Cloud
- Instructions on how to easily integrate existing storage (on-prem, DropBox, etc.) for Cloud usage

Certain details relevant to researchers may include confidential information about vendor contracts and should therefore be HarvardKey-protected where necessary.

4.1.3. Expected outcomes

It is hoped that better communication of available services will attract researchers currently operating outside of University Cloud contracts to join in to central services, which will both better protect the University and also potentially lead to better volume discounts.

4.2. Cloud account and tool provisioning

Implementation of this recommendation generally involves activities that reduce the amount of up-front work that researchers will need to take on to begin using the Cloud.

4.2.1. Required changes

We generally expect that the creation of centralized account and tool provisioning can be achieved using current staffing levels, though it will require some refocusing of current staff. However, it is less clear whether local schools are appropriately staffed to support the researchers once they are onboarded (see 4.3).

4.2.2. Implementation approach

Specific recommendations for HUIT action include:

- For each of the three major providers (AWS, Azure, GCP), use account vending and tools like AWS Control Tower or Azure Secure Enclave to create “landing zones” for researchers to create Cloud resources in a secure, compliant space (note: this will likely require a larger amount of effort in GCP simply because there is less existing GCP experience on the administrative side to leverage for researcher support)

Research in the Cloud

- Investigate central implementation/support for tools like AWS Service Workbench or Azure CycleCloud to reduce the need for researchers to manually provision their own compute resources. If such tools do not offer sufficient service, work with Cloud vendors to modify or build tools such that our researchers' use cases are better met.
- Establish partnerships with, and a community of, local Research Computing groups who would be responsible for onboarding researchers.
- Create a "Science DMZ" (<https://fasterdata.es.net/science-dmz/>) version of HUIT's existing Cloud Shield (aka., the common network and security tools used in administrative Cloud computing) that can provide a base level of security for researchers without significant financial or procedural overhead. This DMZ would need to be able to potentially support L3 and L4 data.
- Investigate requirements for ongoing Cloud hosting of research data and/or research web sites once grants and projects expire, including allowances for reproducibility work by other researchers.

4.2.3. Expected outcomes

Offerings produced under this recommendation will need to be designed to serve the largest number of research needs without overly-burdening existing teams. Hopefully, it will be possible to delegate certain functions such that distributed Research Computing groups could produce templates that could be integrated for larger use. It is also critically important that whatever is produced be perceived by researchers as something that assists them in their Cloud work, not something that gets in their way. If what is produced is not seen as being valuable by researchers, the effort put into it could be wasted.

4.3. Expansion of Cloud research services

Minimally, groups currently supporting research in the Cloud should be in regular communication, which will at least improve existing support levels. Action is already taking place in this direction. However, if Harvard is to truly leverage what Cloud computing brings to research (including more cost effective scaling and utilizing specialized services from Cloud providers), we recommend that the local teams that support on-premises research expand their portfolios to explicitly include supporting Cloud research (whether Cloud-native or hybrid), and that they work closely with HUIT to deliver those services. HUIT should also increase its Cloud Consulting service's focus on research in the Cloud to better support those local Research Computing teams.

4.3.1. Required changes

This implementation would require augmentation of existing school Research Computing teams and possibly the Cloud Consulting team, which will likely require additional financial investment. Cloud-focused teams are fully-occupied in their support of administrative Cloud computing, and Researched-focused teams are fully-occupied in their support of on-premises research services. In addition to investment in human resources, there will also need to be investment in regular training for interested researchers and for the local staff who support them.

4.3.2. Implementation approach

The expansion of local Research Computing portfolios should provide:

- Broad experience across various Cloud providers and various tools to help find best fit
- Software patterns for performing certain types of research in the most cost-effective way possible
- Direct support in building research environments when an existing pattern cannot be found

HUIT Cloud Consulting should:

- Work with Cloud providers to help centralize provider credit visibility and management, to ensure that local IT teams are aware of what credits their researchers are using

Research in the Cloud

- Engage vendors to provide access to their staff for embedding with Harvard to launch these new services
- Create and share cost management best practices for researchers, and document tools that can assist them
- Regularly convene a community of Cloud Research support staff to coordinate activity and gather feedback on central services
- Work with Cloud Research community at Harvard to actively engage with Cloud Research teams at other universities to identify best practices and collaboration opportunities
- Work with vendors, partners like I2 (<https://internet2.edu/cloud/cloud-learning-and-skills-sessions/>), and the IT Academy to make sure that research-focused Cloud training is available for researchers and the local staff who support them

4.3.3. Expected outcomes

It is critically important that this effort be a partnership between HUIT and local Research Computing teams. HUIT can expand its services to provide broad horizontal support of Cloud research, but local Research Computing teams would still be required to provide vertical support.

One major area of concern is whether local Research Computing teams (especially the smaller ones) could reasonably expand their portfolios as described. There may be opportunities to leverage larger local IT teams to augment the capacity of smaller teams, and/or to create shared services or service infrastructures that would be available across Harvard and be leveraged by these smaller teams. This could also help with the problem of scale, where some research projects may require significantly more support than others.

4.3.4. Risk of non-implementation

The consequences of not expanding services to include support of research in the Cloud all derive from the fact that, even without that support, researchers are already going to the Cloud. The potential consequences of this include:

- Inefficient use of financial and human resources as research teams have to bring themselves up to speed on the Cloud before they can begin their research
- Waste of funds when researchers are unable to take advantage of Cloud discounts that they are unaware of
- Prematurely exhausted credits when resources are built without full understanding of Cloud cost models, resulting in less discovery and more effort by financial and technical staff to recoup those credits
- Lack of visibility into total Cloud spend (paid as well as credited) at Harvard, which impacts Harvard's ability to negotiate with vendors (e.g., bulk discounts)
- Security risks if Cloud resources are created without attention to or knowledge of best practices
- No opportunity to redirect researchers to existing on-premises resources that could be more financially beneficial to their research funds and/or to the University as a whole

5. Appendices

5.1. Appendix A: Higher Ed survey

A presentation on the results of the Higher Ed survey can be found [here](#).

5.2. Appendix B: Reasons for researchers to use the Cloud

Researchers are directly going to the Cloud for various reasons despite the availability of on-premises resources. Some of the documented reasons include:

Research in the Cloud

- Cloud services not available in on-premises systems (e.g., specialty Machine Learning tools)
- Ease of scaling Cloud resources up and down to meet need/usage patterns
- Speed of deployment/low barrier to entry in basic Cloud usage
- Experimentation/innovation
- Data sharing/collaboration across institutions
- Federal agency (NIH, NSF) focus on national research platforms in Cloud environments
- Vendor credits from Cloud providers
- Public datasets that are already published to the Cloud

Credits drive many researchers to the Cloud, but may not be a sustainable way to do research in perpetuity. Keeping tabs on credit volumes and credit programs is an area where assistance is already needed by some researchers, and there will likely be more of that in the future.

5.3. Appendix C: Vendor lock-in

It should be noted that the three major Cloud providers are all competitors in a fast-moving field with a constantly developing ecosystem. Vendor lock-in in the Cloud space has long been a consideration for Administrative computing, but it is an issue for researchers as well. Researchers who depend on vendor credits may want to be able to easily shift workloads from one-vendor to another, and (as with administrative computing) the more vendor-specific tools they use, the harder that will be. In addition, any workload that depends on credits needs to have a plan for how the research will operate after the credits run out, which may mean a plan for shifting the workload to a less expensive location.

5.4. Appendix D: Level 4 data

Finally, as we consider the support for Level 4 research data in the Cloud, we will need to balance the need for convenience in putting compute and data close to each other with the security and privacy concerns of working with such data. When researchers find themselves driven to the Cloud because of a lack of ability to host Level 4 data in on-premises research spaces, we should also consider whether some of those use cases could also be filled by a de-identification service that is easy to use and available to researchers across the University.

The risks around Level 4 data in the administrative space tend to focus around the financial risks of penalties and/or payouts, and the reputational risk to the University. The research space, however, has some special areas of concern:

- Varying quantities of data and locations of data subjects can make security risks difficult to quantify. For example, the risk profile of a small study vs. a Big Data study, or the widely differing rules for when study participants are in the U.S. (which state are they in?) vs. in the E.U. (GDPR rules).
- Data lifecycles are more complicated, and can be impacted by data holder terms, funding terms, reproducibility needs, and collaborations with third parties.
- Breaches of research data could result in the loss of intellectual property.
- In addition to the loss of reputation with funding agencies, breaches of the terms of funding resulting from security failures could also result in the loss of the funding itself.

5.5. Appendix E: Student researchers

This report does not specifically address the research needs of students, but such use cases could be taken into account, especially in the space of the account and tool provisioning detailed in Section 4.2.

5.6. Appendix F: Research grants and Cloud funding

Funding agencies like the NIH and the NSF have been encouraging the use of Cloud for researchers using their funding. During early phases of the NIH's STRIDES program, it sounded as though the

Research in the Cloud

funding dollars might go directly from the NIH to the Cloud providers. This possibility raised concerns within Harvard that we might lose out on our ability to assess overhead Facilities and Administrative (F&A) costs to the NIH funding if the funding never hit Harvard's books. While the STRIDES program did not end up distributing funding directly to the Cloud providers, it is important that Harvard continue to monitor the evolution of Cloud programs at funding agencies to make sure none of them head in the direction of funding Cloud providers directly. The ability to assess F&A costs to grants is a critical funding source for the support of Research Computing generally at the University.

5.7. Appendix G: 2020 vendor credits

It is difficult to track down total value of vendor credits and how many of those credits are distributed for research activity in particular. AWS has informed us that, in CY2020, Harvard consumed \$516k in credits, \$250k of which can be attributed to Research and Machine Learning credits. Google has not provided us a similar analysis, but the billing console suggests that in CY2020, Harvard received \$931k in "promotions" that went to HMS (VirtualFlow project) and HSPH (Lin Lab).

6. References

- Research Computing Support by School/Unit - <https://researchsupport.harvard.edu/get-help>
- Harvard's Data Security Policy - <https://policy.security.harvard.edu/view-data-security-level>
- Harvard University IT - <https://huit.harvard.edu/pages/about>