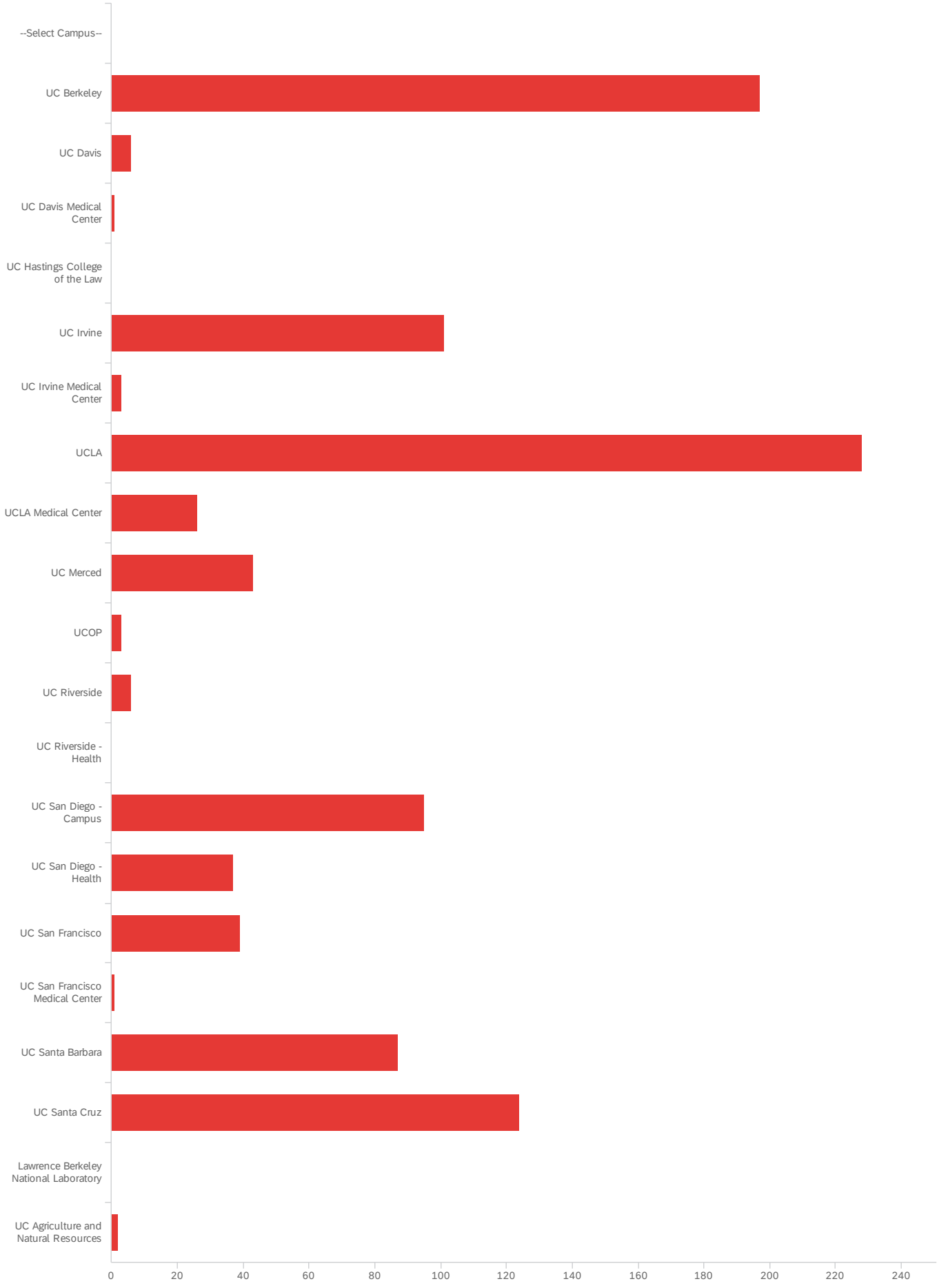
Q2.4 - Campus

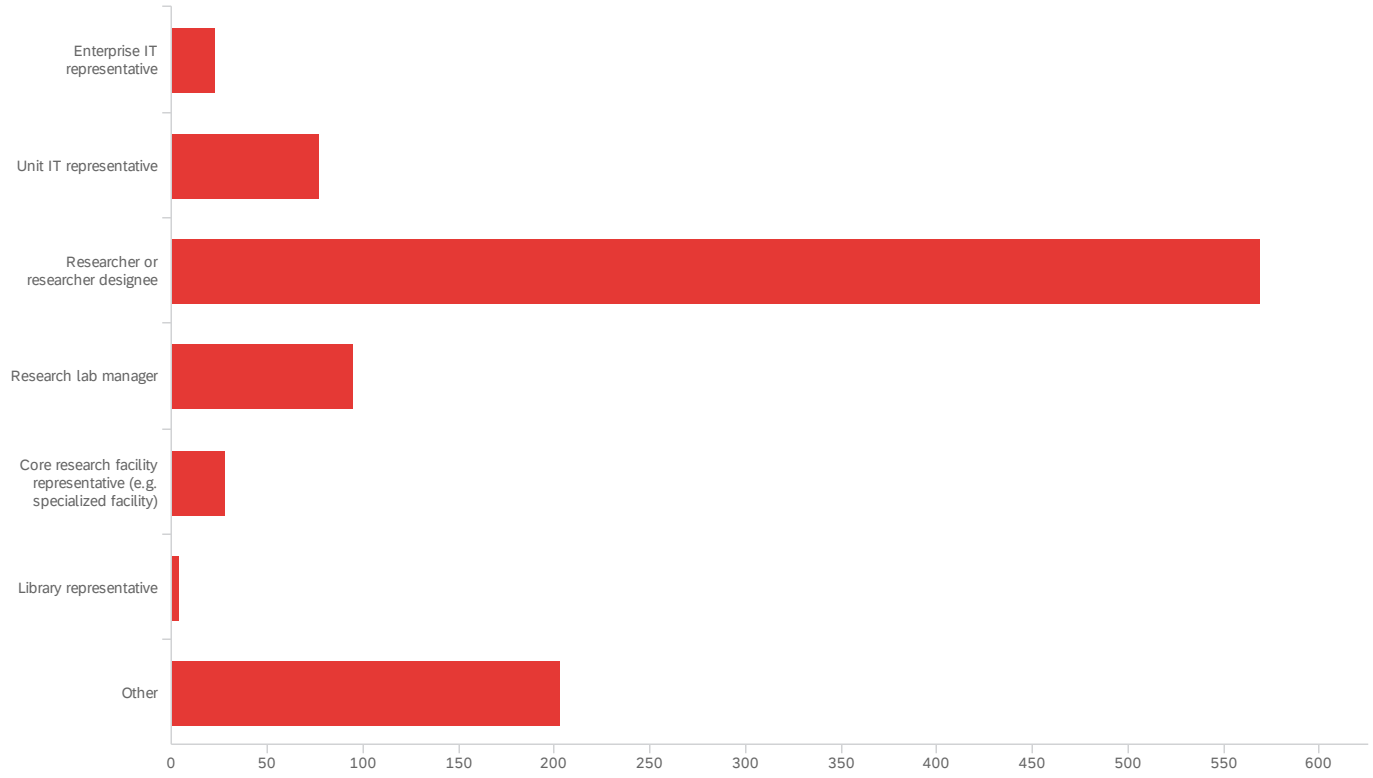

--Select Campus--
UC Berkeley
UC Davis
UC Davis Medical Center
UC Hastings College of the Law
UC Irvine
UC Irvine Medical Center
UCLA
UCLA Medical Center
UC Merced
UCOP
UC Riverside
UC Riverside - Health
UC San Diego - Campus
UC San Diego - Health
UC San Francisco
UC San Francisco Medical Center
UC Santa Barbara
UC Santa Cruz
Lawrence Berkeley National Laboratory
UC Agriculture and Natural Resources
0
20
40
60
80
100
120
140
160
180
200
220
240

#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	Campus	2.00	21.00	10.14	5.94	35.31	999

#	Field		Choice Count
1	--Select Campus--	0.00%	0
2	UC Berkeley	19.72%	197
3	UC Davis	0.60%	6
4	UC Davis Medical Center	0.10%	1
5	UC Hastings College of the Law	0.00%	0
6	UC Irvine	10.11%	101
7	UC Irvine Medical Center	0.30%	3
8	UCLA	22.82%	228
9	UCLA Medical Center	2.60%	26
10	UC Merced	4.30%	43
11	UCOP	0.30%	3
12	UC Riverside	0.60%	6
13	UC Riverside - Health	0.00%	0
14	UC San Diego - Campus	9.51%	95
15	UC San Diego - Health	3.70%	37
16	UC San Francisco	3.90%	39
17	UC San Francisco Medical Center	0.10%	1
18	UC Santa Barbara	8.71%	87
19	UC Santa Cruz	12.41%	124
20	Lawrence Berkeley National Laboratory	0.00%	0
21	UC Agriculture and Natural Resources	0.20%	2
			999

Showing rows 1 - 22 of 22

Q2.6 - Role

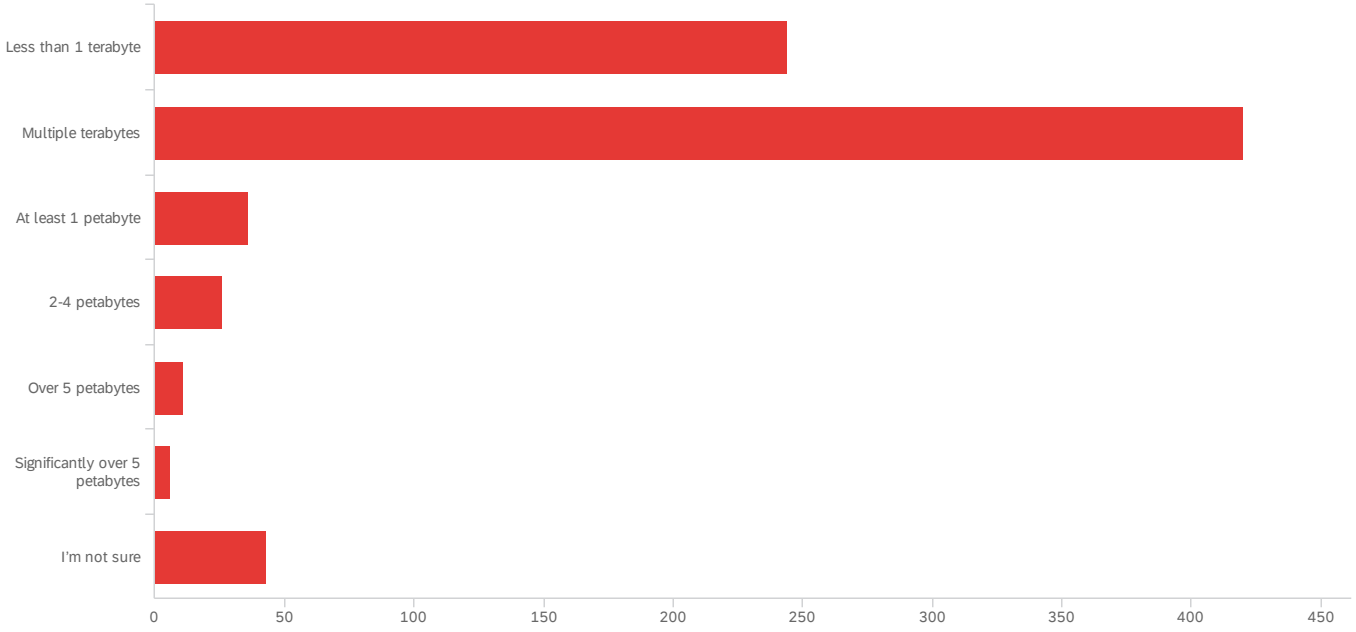


#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	Role - Selected Choice	1.00	7.00	3.85	1.71	2.94	999

#	Field	Choice Count	
1	Enterprise IT representative	2.30%	23
2	Unit IT representative	7.71%	77
3	Researcher or researcher designee	56.96%	569
4	Research lab manager	9.51%	95
5	Core research facility representative (e.g. specialized facility)	2.80%	28
6	Library representative	0.40%	4
7	Other	20.32%	203
			999

Showing rows 1 - 8 of 8

Q3.1 - Roughly, how much research data resides in systems under your control?



#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	Roughly, how much research data resides in systems under your control?	1.00	7.00	2.15	1.44	2.08	786

#	Field	Choice Count	
1	Less than 1 terabyte	31.04%	244
2	Multiple terabytes	53.44%	420
3	At least 1 petabyte	4.58%	36
4	2-4 petabytes	3.31%	26
5	Over 5 petabytes	1.40%	11
6	Significantly over 5 petabytes	0.76%	6
7	I'm not sure	5.47%	43
			786

Showing rows 1 - 8 of 8

Q3.2 - Roughly, what percentage of research data associated with your unit are stored on:

#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	Unit-controlled network drives (NAS or similar):	0.00	100.00	8.71	24.24	587.44	793
2	Lab-controlled network drives (NAS or similar):	0.00	100.00	11.23	25.88	669.88	793
3	Enterprise-controlled network drives:	0.00	100.00	4.98	18.35	336.68	793
4	Desktops/Laptops:	0.00	100.00	20.73	29.10	846.66	793
5	USB drives and other external storage drives and devices:	0.00	100.00	10.23	21.70	470.98	793
6	Enterprise cloud-based shared drives (e.g., Box, OneDrive, Google Drive):	0.00	180.00	22.12	31.64	1000.94	793
7	Clusters, whether a campus cluster or unit-based cluster:	0.00	100.00	7.58	22.12	489.21	793
8	Institutionally provided cloud accounts (e.g., AWS, Azure, GCP):	0.00	100.00	3.94	15.67	245.41	793
9	Personal devices or personal cloud accounts controlled by the individual researchers:	0.00	100.00	6.94	19.42	377.18	793
10	Other (please specify):	-180.00	100.00	3.54	18.36	336.95	793

Q3.2_10_TEXT - Other (please specify):

Other (please specify):
DVDs
AD server
an SQL server, I don't know which of these it falls under
Physical materials
AWS not provided by the institution
BIC server (fMRI raw data)
dropbox

Other (please specify):
don't know the answers to these
UCSD Google Drive
notebooks
the ultimate backup is on hard copy field data sheets
REDCap
national supercomputers
n/a
I have no research data
do not know
VA Enterprise Servers
Stored with Collaborator
bad question since it does not allow redundancy
on-prem backup servers
Unknown
0
Github
Unknown: PIs are responsible for data generated by their experiments at my facility. Facility management data is maintained on college-controlled servers
The same data is stored in several places... so percentagescon't really make sense...
Unit maintained Linux Server
I am only counting the data not on offsite tape storage here.
Not sure
SRDC VM
total to 100 is BS as we have redundancy
The same data is on multiple sources so this question is poorly phrased

Other (please specify):

XSEDE Archival Storage (Ranch at TACC)

I don't have even a rough idea

CommVault back-up system

Health's M365 instance

I use large images

I have everything on my laptop, my desktop and dropbox. That adds up to 300% but that is accurate as it is duplicated across those three platforms.

Non-networked desktop (our MOU requires it)

External data company

other institutions' servers

local SQL Server Machines

do not know

paper files in my office

Paper copies

confusing question - what about duplicate copies? What about Google Drive, it is a cloud account AND drive; next - what is the difference between a network drive and server?

Not sure

google drive

everything is also backed up on UCLA servers periodically, but it's a time-consuming task

duplicates of data are stored at user facilities where data is obtained

Our preferred lab backup solution is now Wasabi.

Compensation for offsite backups

A mix of all of the above

laptop and dropbox (redundant)

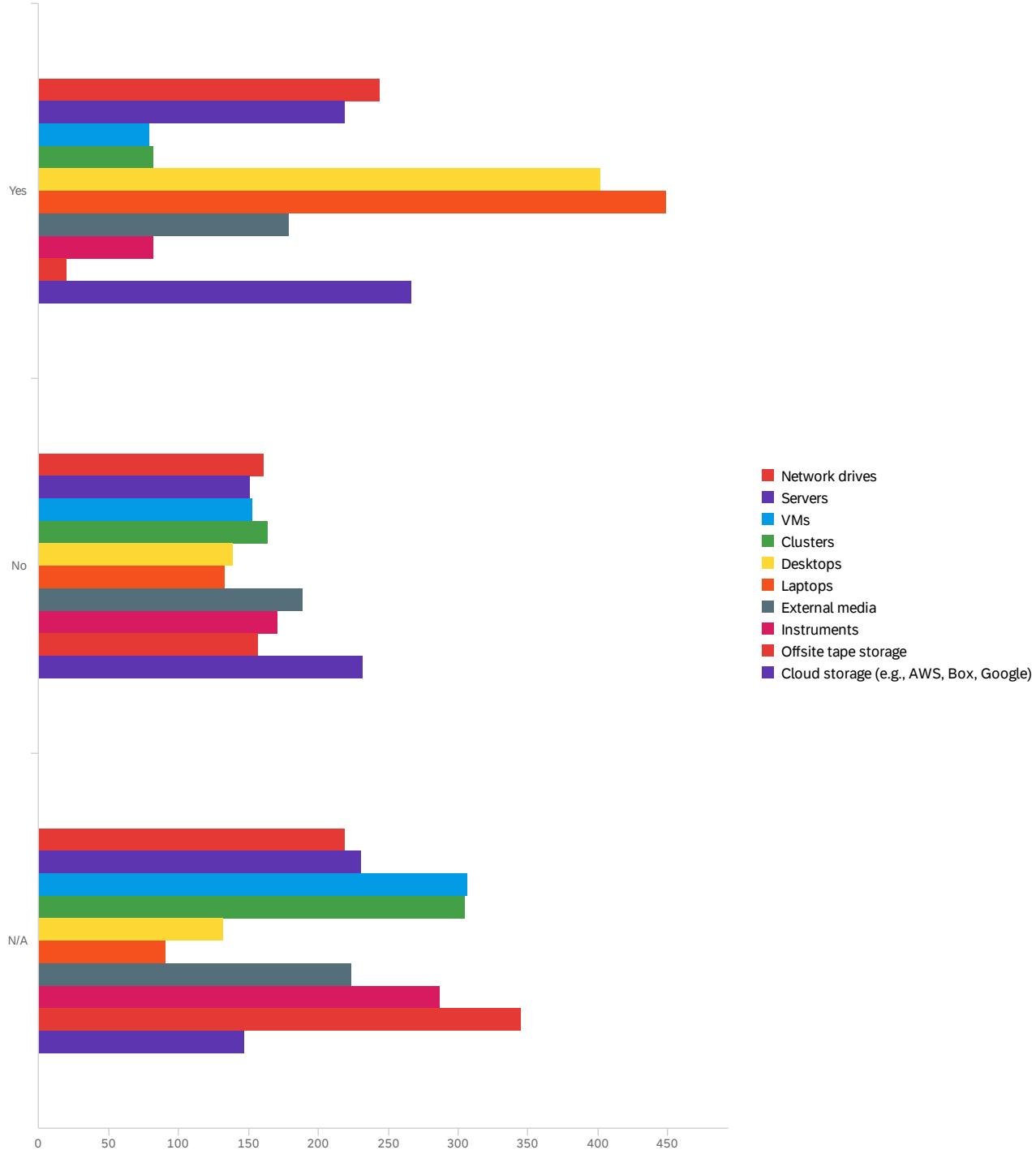
UCI Qualtrics 50%, Dropbox 50% (de-identified data)

UCSD Sherlock

Other (please specify):

Dropbox
Compute servers
I work independently of any unit
uci server
Instruments
tape

Q3.3 - Current BU



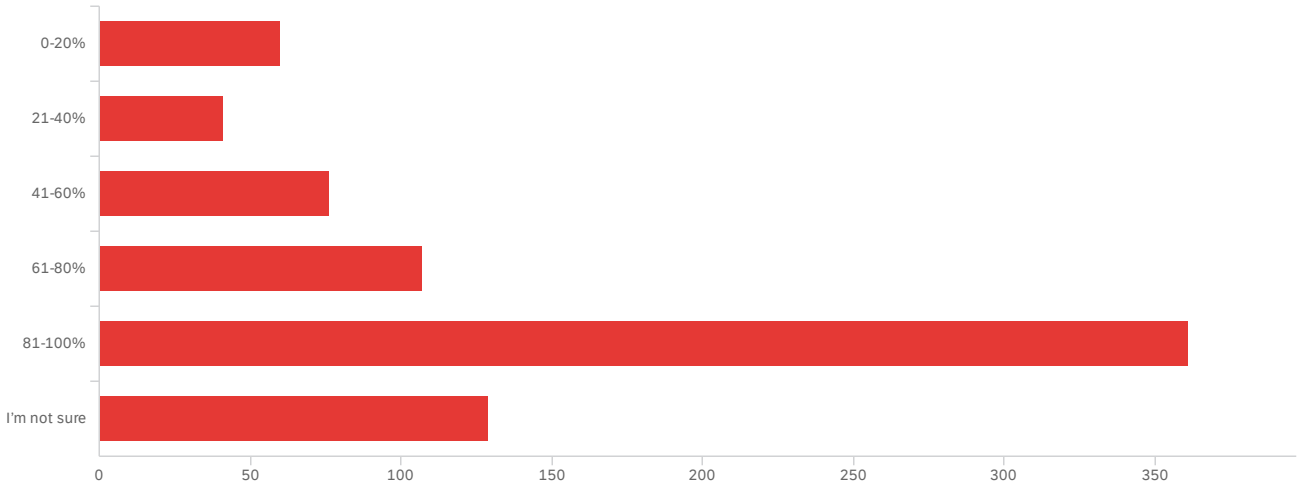
#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	Network drives	1.00	3.00	1.96	0.86	0.74	624
2	Servers	1.00	3.00	2.02	0.87	0.75	601

#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
3	VMs	1.00	3.00	2.42	0.73	0.54	539
4	Clusters	1.00	3.00	2.40	0.73	0.54	551
5	Desktops	1.00	3.00	1.60	0.80	0.63	673
6	Laptops	1.00	3.00	1.47	0.72	0.52	673
7	External media	1.00	3.00	2.08	0.82	0.67	592
8	Instruments	1.00	3.00	2.38	0.73	0.54	540
9	Offsite tape storage	1.00	3.00	2.62	0.56	0.31	522
10	Cloud storage (e.g., AWS, Box, Google)	1.00	3.00	1.81	0.78	0.61	646

#	Field	Yes		No		N/A		Total
1	Network drives	39.10%	244	25.80%	161	35.10%	219	624
2	Servers	36.44%	219	25.12%	151	38.44%	231	601
3	VMs	14.66%	79	28.39%	153	56.96%	307	539
4	Clusters	14.88%	82	29.76%	164	55.35%	305	551
5	Desktops	59.73%	402	20.65%	139	19.61%	132	673
6	Laptops	66.72%	449	19.76%	133	13.52%	91	673
7	External media	30.24%	179	31.93%	189	37.84%	224	592
8	Instruments	15.19%	82	31.67%	171	53.15%	287	540
9	Offsite tape storage	3.83%	20	30.08%	157	66.09%	345	522
10	Cloud storage (e.g., AWS, Box, Google)	41.33%	267	35.91%	232	22.76%	147	646

Showing rows 1 - 10 of 10

Q3.4 - What percentage of the research data associated with your unit is being backed up?

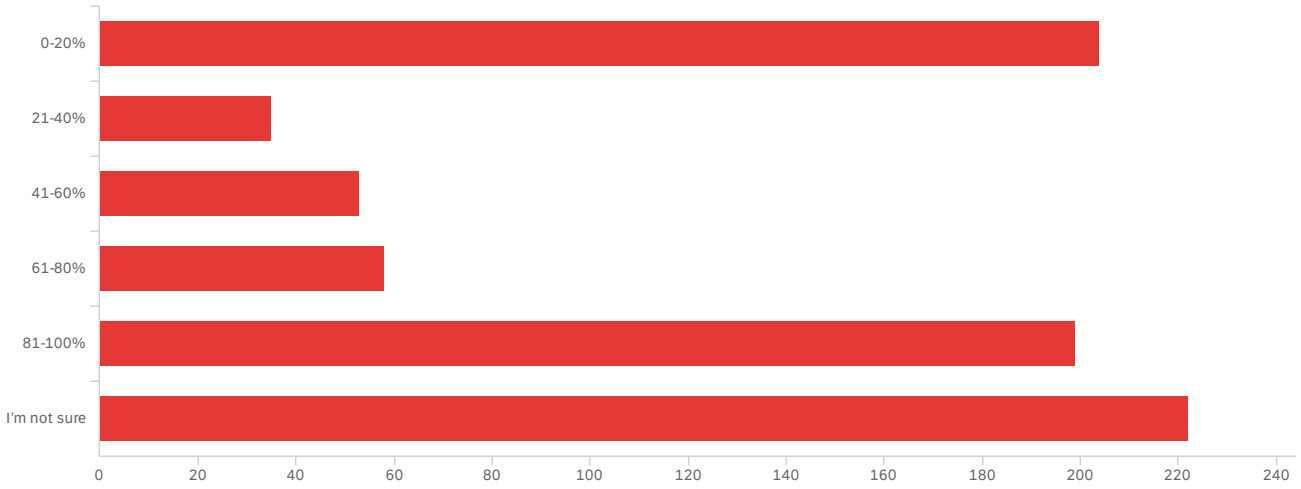


#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	What percentage of the research data associated with your unit is being backed up?	1.00	6.00	4.36	1.42	2.01	774

#	Field		Choice Count
1	0-20%	7.75%	60
2	21-40%	5.30%	41
3	41-60%	9.82%	76
4	61-80%	13.82%	107
5	81-100%	46.64%	361
6	I'm not sure	16.67%	129
			774

Showing rows 1 - 7 of 7

Q3.5 - What percentage of the research data that is being backed up is also covered by off-site backups?

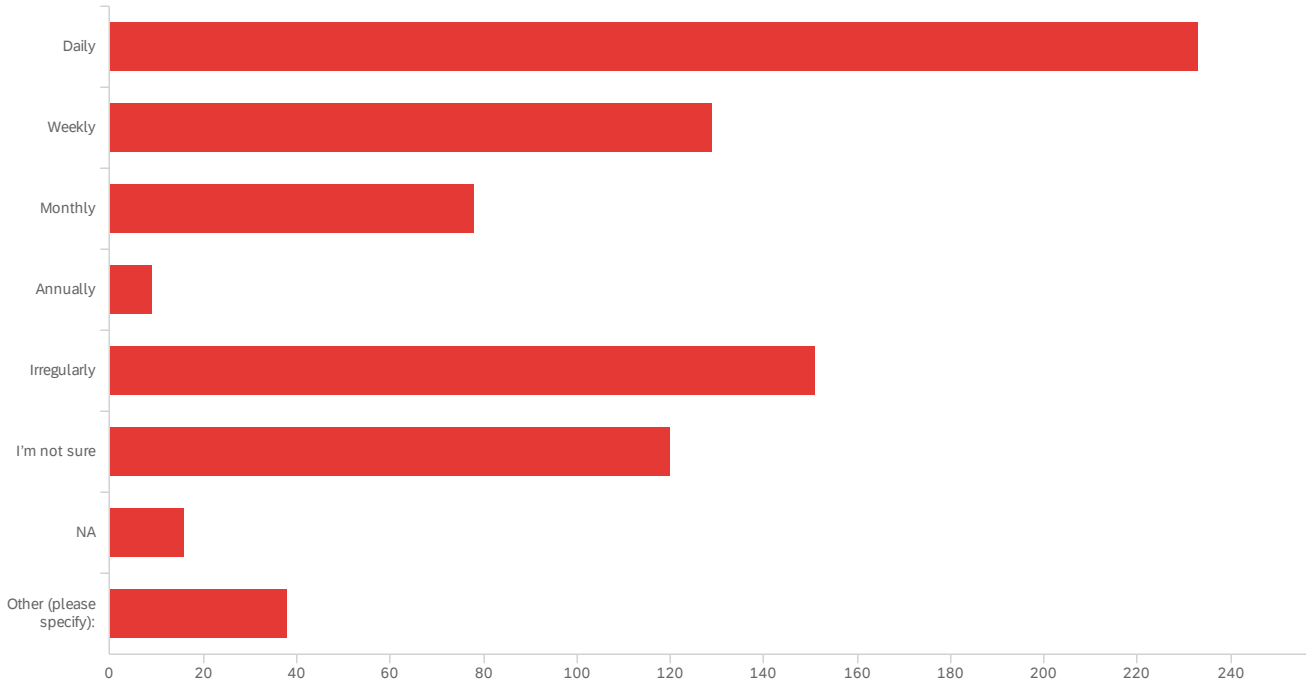


#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	What percentage of the research data that is being backed up is also covered by off-site backups?	1.00	6.00	3.88	2.01	4.03	771

#	Field	Choice Count
1	0-20%	26.46% 204
2	21-40%	4.54% 35
3	41-60%	6.87% 53
4	61-80%	7.52% 58
5	81-100%	25.81% 199
6	I'm not sure	28.79% 222
		771

Showing rows 1 - 7 of 7

Q3.6 - On average, how often is the research data under your control backed up? (Select one)



#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	On average, how often is the research data under your control backed up? (Select one) - Selected Choice	1.00	8.00	3.43	2.22	4.93	774

#	Field	Choice Count
1	Daily	30.10% 233
2	Weekly	16.67% 129
3	Monthly	10.08% 78
4	Annually	1.16% 9
5	Irregularly	19.51% 151
6	I'm not sure	15.50% 120
7	NA	2.07% 16
8	Other (please specify):	4.91% 38

#	Field	Choice Count
		774

Showing rows 1 - 9 of 9

Q3.6_8_TEXT - Other (please specify):

Other (please specify):
Just whatever backing up Google Drive does
daily snapshots
Once (the data is not regularly being created; it was fixed and backed up)
Raw Offline backups are one offs(lab copy & offsite copy). Any actively processed data is backed up daily.
Hourly
Every 5-10 years
Every 1-2 months on external drive, daily for cloud-based
Hourly for 10-day running window and daily for 1-year running window.
Replicate code repositories multiple times a day, git on local server and github. Monthly for data on desktops.
Clusters and servers daily, laptops probably weekly or less frequently
Cloud data I guess daily by Google; laptop data monthly.
Upon collection
Facility management data is backed up daily with UC Backup
I use cloud storage (with version control); I assume this is a form of "backup" already.
unless individuals in my group, e.g. myself, back up their laptop (I do it in dropbox, and pay for it myself personally), there is no backup. Source code is backed up in GitHub. Some group generated data is kept in 2 copies on two different storage solutions. We are transitioning to CEPH with erasure encoding. We currently use HDFS with replica=2.
Once, at the time of download
It use to backup daily but PBSci its told us that they are over the limit and can not backup anymore
Most is on UCSF-controlled systems, which I assume are backed up frequently.
Real time
depends on the data

Other (please specify):

never. no where to do it on site or funds to cover the cloud costs.

the server is a Raid10 with NAS

All data on cluster is auto backed up (not sure if off-site); everything else irregular/set by each individual

DGIT (David Geffen IT) does all Back.

laptop data--daily. other data--less often

My lab server is a set of 6 hard-drives that are mirrored (Raid-5, backed up by the minute), but unfortunately this server is not backed up elsewhere.

it depends on the data set update frequency

Whenever changes/additions to the data archive are made

Instantly

New data is backed up daily. Old data is backed up irregularly

Only as much as cloud is backed up

As changes are made

per cruise

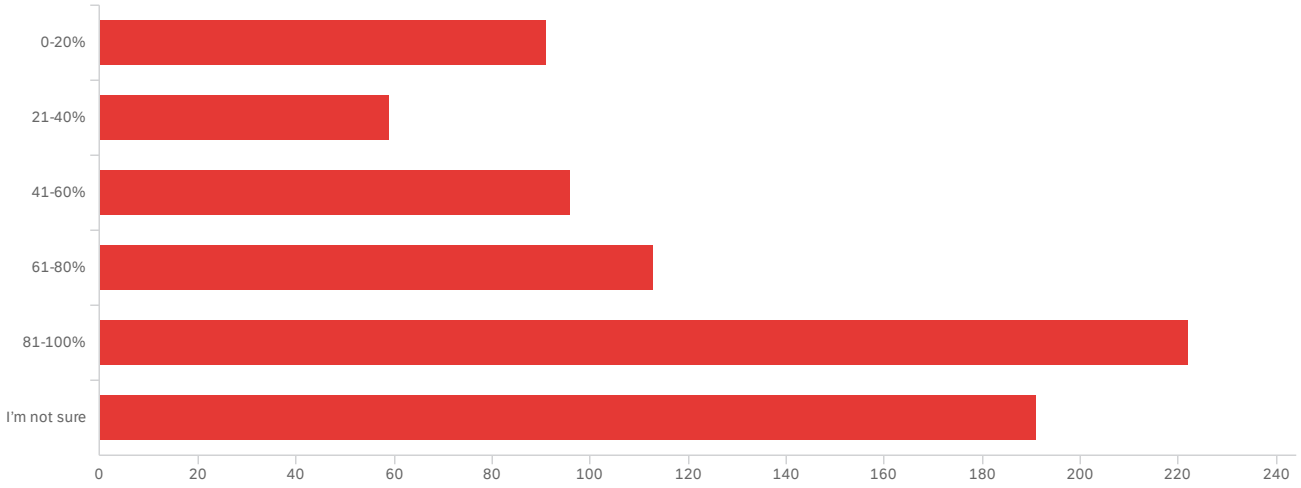
RAID 1 and 50

quarterly

Depends on the data. Some are in regular use, backed up daily, other data is not.

my lab runs/manages a vital records management system for CDPH. not research, but is data managed with UC resources under contract. We have point-in-time recovery

Q3.7 - If the research data associated with your unit were to become unavailable (e.g., as a result of a ransomware attack), roughly what percentage do you think could be easily restored?

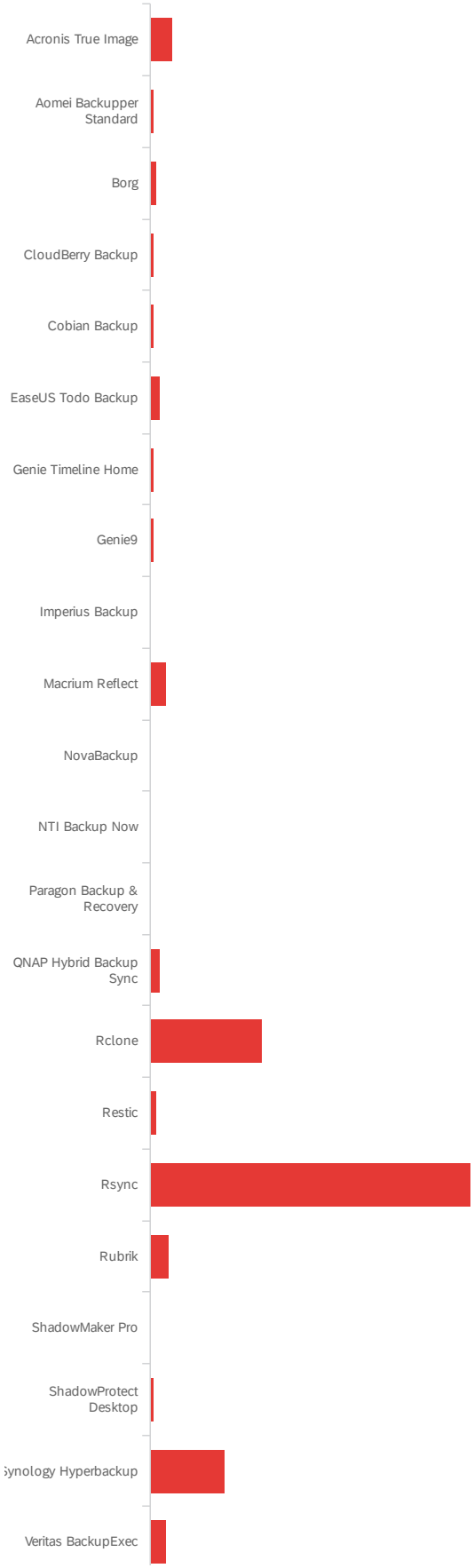


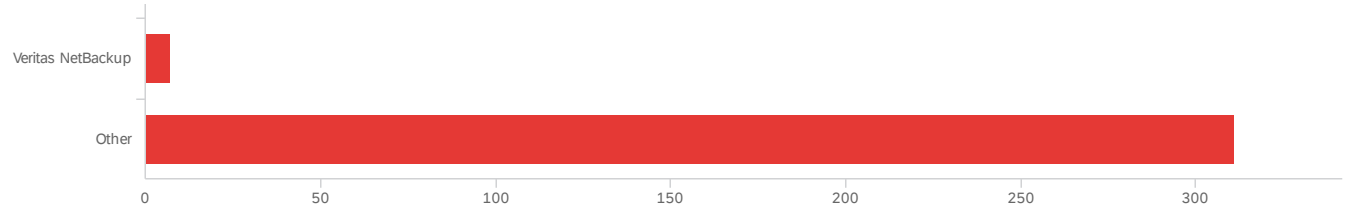
#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	If the research data associated with your unit were to become unavailable (e.g., as a result of a ransomware attack), roughly what percentage do you think could be easily restored?	1.00	6.00	4.15	1.66	2.75	772

#	Field		Choice Count
1	0-20%	11.79%	91
2	21-40%	7.64%	59
3	41-60%	12.44%	96
4	61-80%	14.64%	113
5	81-100%	28.76%	222
6	I'm not sure	24.74%	191
			772

Showing rows 1 - 7 of 7

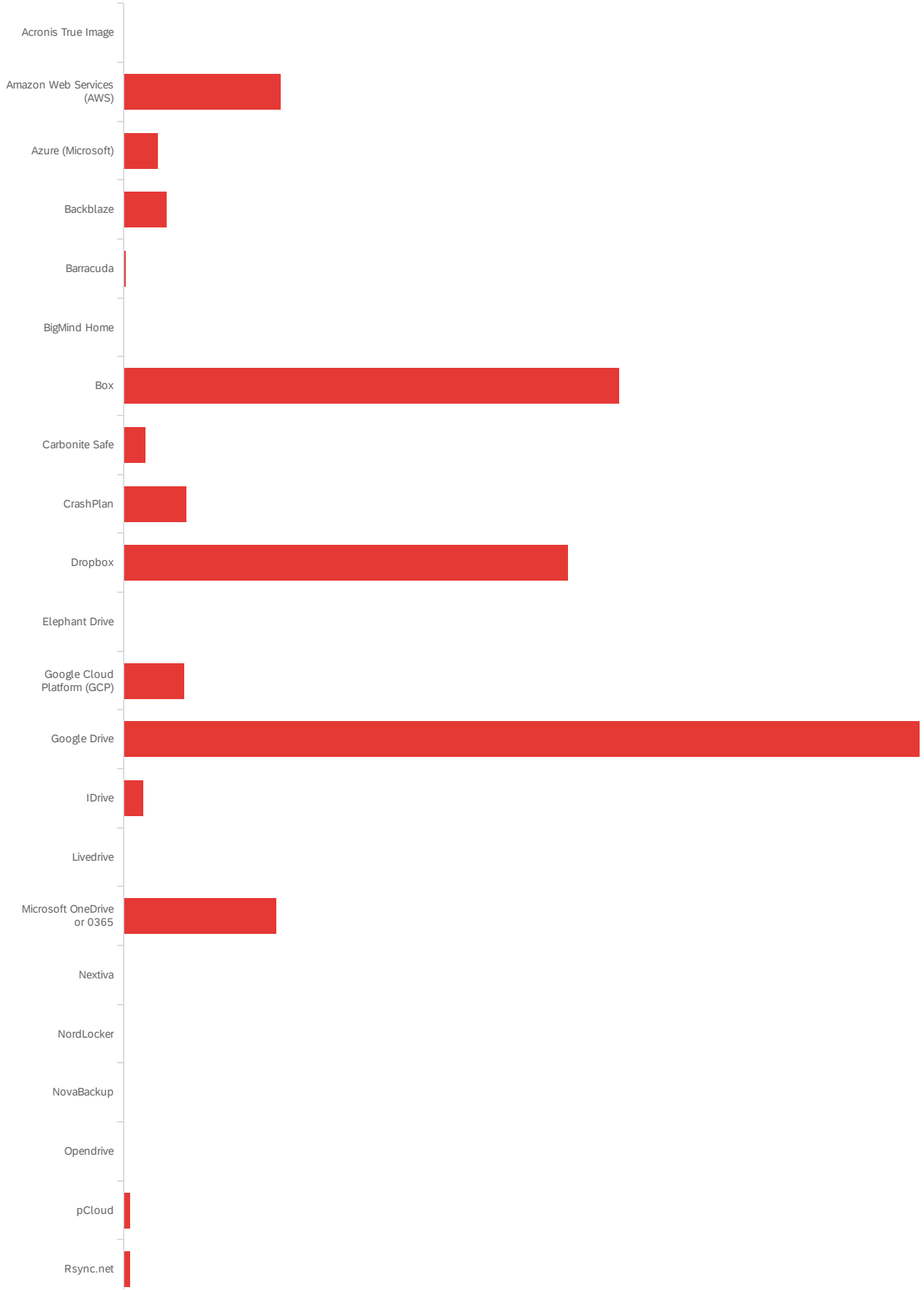
Q4.2 - If on-premises hardware/software, what are you using? (Select all that apply)

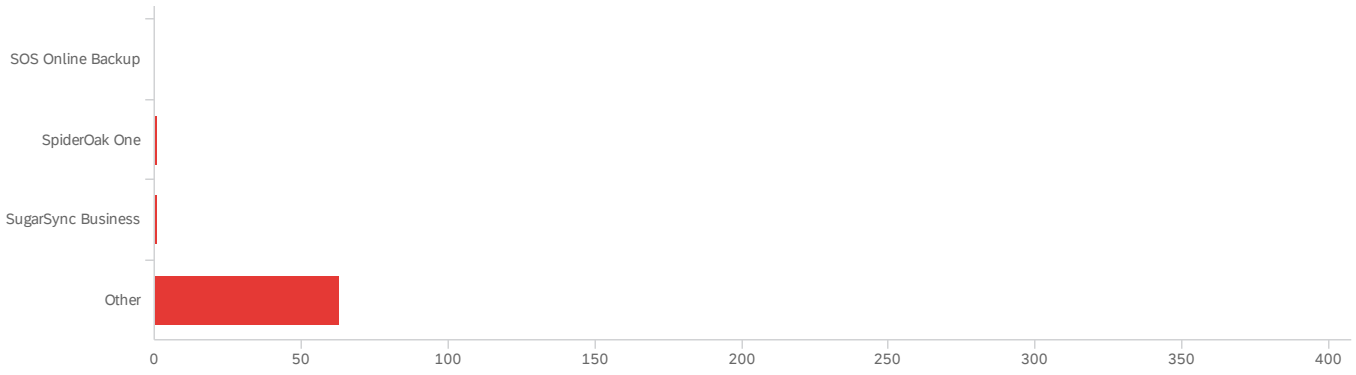




#	Field	Choice Count	
26	Acronis True Image	1.35%	7
27	Aomei Backupper Standard	0.19%	1
28	Borg	0.38%	2
29	CloudBerry Backup	0.19%	1
30	Cobian Backup	0.19%	1
31	EaseUS Todo Backup	0.58%	3
32	Genie Timeline Home	0.19%	1
33	Genie9	0.19%	1
34	Imperius Backup	0.00%	0
35	Macrium Reflect	0.96%	5
36	NovaBackup	0.00%	0
37	NTI Backup Now	0.00%	0
38	Paragon Backup & Recovery	0.00%	0
39	QNAP Hybrid Backup Sync	0.58%	3
40	Rclone	6.92%	36
41	Restic	0.38%	2
42	Rsync	19.81%	103
43	Rubrik	1.15%	6
44	ShadowMaker Pro	0.00%	0
45	ShadowProtect Desktop	0.19%	1
46	Synology Hyperbackup	4.62%	24
47	Veritas BackupExec	0.96%	5
48	Veritas NetBackup	1.35%	7
49	Other	59.81%	311
			520

Q4.3 - If a cloud service, what are you using? (Select all that apply)



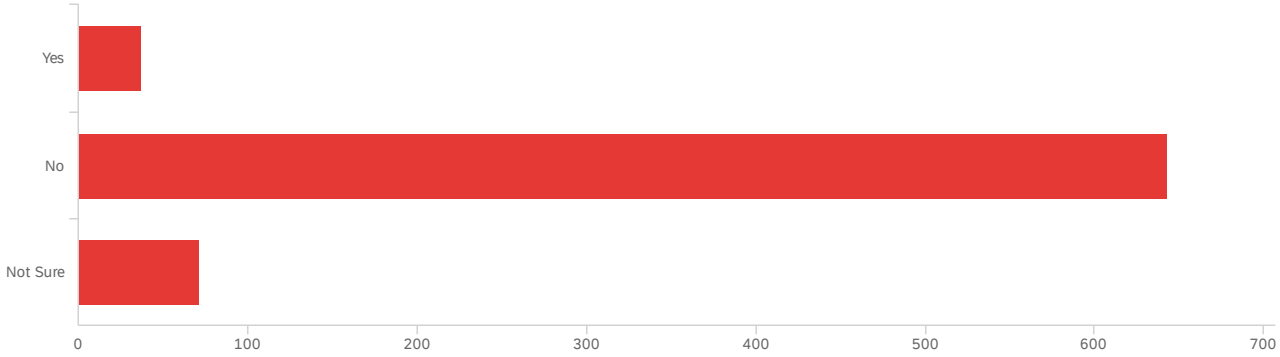


#	Field	Choice Count
1	Acronis True Image	0.00% 0
2	Amazon Web Services (AWS)	6.42% 73
3	Azure (Microsoft)	1.41% 16
4	Backblaze	1.76% 20
5	Barracuda	0.09% 1
6	BigMind Home	0.00% 0
7	Box	20.32% 231
8	Carbonite Safe	0.88% 10
9	CrashPlan	2.55% 29
10	Dropbox	18.21% 207
11	Elephant Drive	0.00% 0
12	Google Cloud Platform (GCP)	2.46% 28
13	Google Drive	32.63% 371
14	IDrive	0.79% 9
15	Livedrive	0.00% 0
16	Microsoft OneDrive or 0365	6.24% 71
17	Nextiva	0.00% 0
18	NordLocker	0.00% 0
19	NovaBackup	0.00% 0
20	Opendrive	0.00% 0
21	pCloud	0.26% 3
22	Rsync.net	0.26% 3

#	Field	Choice Count	
23	SOS Online Backup	0.00%	0
24	SpiderOak One	0.09%	1
25	SugarSync Business	0.09%	1
26	Other	5.54%	63
			1137

Showing rows 1 - 27 of 27

Q4.4 - tape bu



#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	tape bu	1.00	3.00	2.05	0.38	0.14	751

#	Field	Choice Count
1	Yes	4.93% 37
2	No	85.62% 643
3	Not Sure	9.45% 71
		751

Showing rows 1 - 4 of 4

Q4.5 - Other tools, strategies, comments, or clarifications regarding responses to earlier questions:

Other tools, strategies, comments, or clarifications regarding responses to...
Mac time machine
We hand copy things onto external hard drives or painfully, slowly, ever so slowly, upload them into Box.
Not sure what software is being used for backups.
I utilize a multi-stage backup operation using multiple servers with redundant disks in multiple buildings. With my non-destructive backup setup, a ransomware attack might (though unlikely) get the primary data, but would not be able to compromise my 2nd or 3rd level. That being said, my level of paranoia would gladly appreciate another level from the University.
All our research documents are saved on UCB's Google Drive. We rely on its backup system, and don't have our own backup methods.
some data is backed up on DVDs
I assume that the university is backing up the AD server but would like to know, if not as it would be very difficult for us to restore our data if it disappeared.
Utilizing snapshots but not a backup solution; some of the data is already a secondary copy
We recently purchased new instrumentation that will soon be collecting terabytes of data per day, and we have no plan in place for where to save it. Our grant application had a letter in it saying that the university would provide a solution, but thus far, it has not materialized.
"rsync -ahP [source] [destination] 2> ~/Desktop/rSyncErrors.txt" works really well.
I would like to backup 100% of my data weekly at the minimum. I currently keep 90% on external hard drives, but would like to migrate it to storage at the CoLo facility.
I backup to a TimeMachine hard drive.
All of our data is backed up by IGPP NetOps.
Backup to disk and then to tape. Veeam is used to backup VMware virtual servers.
Our laptops are controlled by UCSC and presumably they are backing things up. They're supposed to be doing that..... I back up individual files on an external hard drive.
All primary research data is stored on SIO-managed network drive, and backed up daily to cloud. Working files on individual desktops/laptops are backed up daily to external drives.
My Enterprise-controlled network backup is through SIO IT. I don't know what software is used, but the backup of the networked disk is a week of daily snapshots kept on the server and 90 days worth of daily snapshots on the backup server.
All backups are either on premise disk based backup or cloud.

Other tools, strategies, comments, or clarifications regarding responses to...

Github

macOS Time Machine for office desktops in addition to crashplan

Windows System Image backups occur monthly. Windows backups occur Weekly. Actively rocessed data backed up daily to external drives. All long term non-active and raw data backups have a copy on site and off site.

All our data is on a lab server with a backup on another server. We are moving our primary file server to Wynton, but will continue with a lab-managed server as our backup.

Vast majority of research data is stored on ZFS. Labs are generally setup with a dual server solution, primary server and backup server. Backups are then doing using ZFS snapshots, which tend to be synced hourly from primary to backup.

We use UCSD Druva to back up all of our reading center review stations and staff laptops.

Most of our data is on a server at the Warren Hall co-location facility, but I do not know the backup schedule or technology. This system is overseen by Joyce Gross and Michelle Koo for the Berkeley Natural History Museums. I manage a subset of actively used research data with a DropBox account.

These questions don't really make sense to me. I recently purchased two systems for my research group, both rack mounted: SambaShare server and backup system. Both are set up as RAID 6 with 60 and 95 Tb of usable space. These are for use by my research group (10 people), working with Mac, PC, and Linux workstations. These machines cost ~20k total, had to rise funds from multiple research grants to cover the cost.

We rely on Box and Google Drive and the Savio Cluster for all our data because they are backed up.

I have no idea what 'solution' is being used for the backups, since it is handled entirely by the IT group at Scripps. My 'unit' is just my small lab group, consisting of myself, 1 part-time programmer and 2 PhD students.

New here so answers to these questions don't yet apply. At my prior institution there was a server each lab had an account on that could store large amounts of research data. This could be accessed on campus or by VPN, and was automatically backed up to Amazon Glacier or similar services behind the scenes. I have not found any similar service at Berkeley but believe this should be provided.

I personally tend to save all of my data directly to the Cloud (used to be Box before the 50 GB quota was instituted) as well as on an external hard drive (backed up less frequently). We transfer files from some lab/instrument computers via USB or upload to the cloud, as is reasonable. Some large simulations / files are frequently not backed up due to storage limitations. After almost losing all of our data from a shared lab computer when the computer crashed, we have now installed Backblaze onto lab computers that are remotely accessed most frequently.

All of our raw data and project working directories are stored on a Synology NAS. Occasionally, some data will be copied to a linux workstation during analysis, but is copied back to the NAS asap. The NAS is synced to SDSC USS storage purchased through Research IT as well as incremental backups to AWS Glacier using Synology Glacier Backup.

Our research group uses a shared Box drive for our work, owned by an SPA account. We back that up to a local (at the university, in our lab) Synology 4 bay NAS DiskStation, RAID 1. There are automated weekly differential backups and also a few independent 'Full' backups at 3 month intervals. We would need to lose access to both Box and the NAS for us to lose access to the data.

Primarily using ZFS with snapshots and replication. Most desktops use BackupPC or Duplicati

We have all of our data on google drive, as a lab policy because ITS told me this would be the best set up for the lab. I have no heard that we might end our contract with unlimited google drive space - this is really concerning and would cause significant problems in the lab (rather, we would need an ITS solution to do backups!).

Other tools, strategies, comments, or clarifications regarding responses to...

I have been backing up on external RAID drives (I have three such drives totaling 81 TB. I just added a 42 TB drive so this will allow me to back up research data that currently resides only on a single RAID drive. If someone were to break into my office and steal my computers and drives, my entire research program would be devastated. Some sort of online accessible backup system would be extraordinarily valuable to me. Currently, there is just too much data on my backup drives to easily transport off-campus to my home, so I am basically just holding my breath and hoping I am not burgled or we don't have a natural disaster that destroys all of my data in one fell swoop.

Our backup is just google drive for some data and laptop backup on Apple Cloud

na

I back up critical data not written in notebooks (i.e. photomicrographs) on an external hard drive.

most of these questions are irrelevant to my situation

Data from projects all stored on REDCap

QEMU Snapshots

"Easy to restore" from your earlier question is relative-- while it would be possible to restore all of our backed up data, it would be extremely time consuming due to the large data volumes and slow network speeds, and thus not "easy".

We backup to local resources using open source tools.

I have a much larger dataset at NSF datacenters. When projects finish, these datasets need to move. Where should I move them?

"Off site" is difficult term. We have our data backed up on a server about 0.5 miles from the main data server. The backup server is "on line" but has no user logins, only an admin login, and there are firewall constraints.

A solution to data back up is critical. We cannot continue to oppertate the way we do with o7r own home grown backup solutions.

Most of my data are on AD.UCSD.EDU drive, which I understand to be backed up regularly but I don't know the details.

We extensively use Box. We have been told there is an upper limit of ~10 TB, which may not be sufficient in some cases. The other problem is that accounts are deleted once you a student or postdoc leaves. A shared repository provided by the UC system could overcome these limitations.

There was no option for data being on lab-controlled servers, so I put "Desktops/laptops". Our data is on a server that doesn't use NAS.

iCloud

I'm paying this out of pocket for the backup of my work data

Backups are stored on RAID6 enterprise disk drives on a file server in the datacenter (physically distanced from the original data by about 10 feet (separate rack cabinet from the original data). Backups are mounted read only on the compute cluster. Incremental backups are performed nightly and stored for 7 days. Fileservers are not accessible from the internet. login access to file servers on private network is restricted to administrator with ssh keys.

We might have a small storage quota at SDSC through the department. However, IT did not respond to an inquiry.

I run my own synology NAS to backup data from virtual machine on campus

Other tools, strategies, comments, or clarifications regarding responses to...
I don't understand how or whether all our coding data on Dedoose is being backed up
My entire group store our data at the USS drive managed by San Diego Super Computer center and we used the methods they recommended for the back-up
I back up daily to a Synology NAS. That NAS gets backed up weekly to AWS. I also backup my laptop and desktop continuously to iDrive.
Data is mainly being backed up in OneDrive through the UCSD system.
Snapshots are used on our NetApp and ZFS based storage systems, both local copies, and remote copies
Using on-prem Code42, formerly CrashPlan Pro. Offsite in our case is a secondary server on a different building.
I'm responding from the Enterprise storage and backup standpoint. I do not have any idea what data is research vs not research. If there exists data not stored on Enterprise storage, I am not aware of it, nor how, or if, it is backed up.
Only data stored on the K-drive are backed up. Data stored in the lab are not backed up.
Backing up entire VM using Veeam and storing images for quick restore or gfs versions off site. Not doing file based backup or tape.
We manage all our backups directly. We cannot backup petabytes of stored data, but we do backup home directories and essential software.
Most (>80%) of our data are automatically backed up to box.
We use 24TB Synology RAID archive
I would love an option for a departmental or university wide backup to server/cloud/offsite; our backup "solutions" are currently in the dark ages of manual/intermittent backup which I intend to change though currently uncertain of the path forward. We are a shared use instrument facility (stable isotope lab) responsible for many people's raw data.
local cloud storage
Due to a change in the Box (and soon google drive) contracts, we are currently changing our backup strategy. We will backup both our NAS and server (85% of our data) to AWS.
Our campus provides a backup service, Crash Plan, but doesn't allow encryption where the key is only held by the researcher. This is a problem for storing sensitive data, and prevents me from using this service.
Nothing grand, nothing large size, nothing special
Dependent on one drive and Dropbox backups
data backed up to external drives
I have a lot of data obtained from other researchers for my group to use. So we are not the original data owner, and in principle the data could be restored. But it would be painstaking. As such, we don't hold the originals so we're less vulnerable
These questions all seem like they're better answered by an IT person. A little hard for me (a researcher) to understand all the language. Also, I am answering for our individual research group, but lots of questions are about our whole unit, which I can't answer for.

Other tools, strategies, comments, or clarifications regarding responses to...

There are two data sets: one is the data associated with managing the facility, which is under my direct control, stored on a college-maintaned server, and is backed up with UC Backup. The other is research data maintained by the individual PIs who conduct their research onsite. I have no control over that data and no information about how it is managed.

my backups and storage solutions are very ad hoc- often external drives for laptops and share Google drives. My division is helping to back up a few hundred gigabytes of research data on some shared lab computers.

This survey was really confusing, and hard to tell what different parts were referring too! "Unit" wasn't defined, and I wasn't sure what was meant by it... my lab? My department?

The researchers who use RCNR's file server have their data backed up with the UC Berkeley campus backup solution which run on Commvault. These are not all RCNR researchers and I cannot speak for the others.

Basically we store our data on Box and/or Sharepoint. We try to regularly back up to Sharepoint. Data are also stored on the cloud account of survey providers, e.g, Qualtrics or RedCap.

Regarding the question about the % of data stored in different ways: this should not be constrained to add up to 100% because. Most of my lab's data are stored in multiple places (e.g., more than one computer, external drives and also with a cloud based service).

We have our data separated in three different places - 1) Data with UCB IT Data Center and this is backed to UCSD 2) Data with our unit controlled HPC is backed up to Amazon Glacier 3) Data with our lab/biorepository is backed up to Amazon Glacier and/or the UCSD

We assume that data stored on RAE and Box is backed up by ARS/IT.

We store data primarily on box and a DGIT controlled network server. We also have local external hard drives that we back-up data to.

I don't understand some of the questions/terms in previous pages

I am not providing ANY back up services. I tell the user of my facility that it is their responsibility to back up and remove their data from my instruments. This is a terrible and inadequate situation.

Please get us Dropbox

I use 100% cloud storage (with version control); I assume this is a form of "backup" already.

The bulk of the 5++ PB of research data we have is backed up at Fermilab. In this report, I am mentioning only the ~1PB that is not backed up that way. In fact, it's probably less than 1PB but we provision up to 1PB of storage for user generated data. In general, all the data that is not backed up on tape at Fermilab is user generated data that users can regenerate IFF they have kept track of their source code in GitHub as they should, and opt provenance for what they did. Some of this is organized at the group level, and some of it is organized at the individual level. Recreating al the user data that is relevant would take us at least a month, if not more. Possibly a couple months.

We expect our researchers to back up their own data by themselves--our servers are for sharing data, not archiving it, so we expect they have their own copies or means to restore their data. I have diligently tried to use the Library's systems to store my research data but was stopped by the requirement for me to personally indemnify the University in case of lawsuits. I cannot put my family home and retirement funds at risk to pay lawyers on the University's behalf while I'm doing my job (and I can't buy insurance to protect myself in this case either). Your mandatory agreement states: "You agree to indemnify UC San Diego against any third party claim based on an allegation that UC San Diego has violated that third party's copyrights, trademarks, other intellectual property, privacy, publicity or other legal rights due to your actions or omissions." So, if this personal indemnification continues to be mandatory, I can't use university backup solutions.

we are not backing up to off-site

I'm not sure how to answer the prior questions, because I'm not sure what you consider to be "under my control" as chair. For example, researchers in my Department have data stored - is this "under my control" since I am chair, or is this under their control (since realistically, I can't control where or how they are storing data or doing backups).

Other tools, strategies, comments, or clarifications regarding responses to...

Use of CommVault for on-prem backup. Looking to move to their SaaS offering.

our unit had research data to the ransomware attack. before the ransomware attach, we used a mix of enterprise-level data storage (UCSF Research Analysis Environment) - 80%- and department-level data storage. We now depend on RAE nearly exclusively for data storage and backup.

I hope the UC system can move to Dropbox or Google Drive for cloud-based active storage and backup, instead of Box.

Majority of lab data is held on a shared google drive. Large datasets are typically stored locally and backed up on additional external hard drives. A complete back up of the google drive data is downloaded at least annually by the professor.

we do not have a uniform backup system for our laboratory computers yet (new lab). I manually back up most of my data at the moment.

Druva Backup

backing up onto multiple external USB drives

What I need, ideally, is a backup option that I can use on Linux from the command line. I currently use my personal Dropbox. The university provides a Google Drive, but I cannot use it in the same way as Dropbox. My DropBox is not large enough to hold all of my data.

Our research data is all in MS SQL Server and currently at about 30GB in size. We maintain daily back ups for a year. The size of the backup dataset is why we selected multi-terabytes of research data.

Using AWS S3 storage with daily versioning turned on.

Does backing up desktop and laptop data to a shared Box drive owned by a campus SPA- account count as a backup? It's redundancy not still vulnerable to deletion.

We use AWS RDS with clustering and backups. We also use AWS S3 with versioning. Our infrastructure configuration is stored off-site (github.com)

We keep instrument and other computers OFF the Net as much as possible.

All data stored in all of the three (ideally, sometimes students skip steps): local computer, cloud storage and external HDD not connected to computer when not in use

We are doing dally mirroring and backup over a period of 5 years

Data management plan requires that I back up research data. But there is now a limit to my Box storage which will eventually prevent me from doing so.

We literally have resorted to a final backup in another building using a hard drive which is carried over once a month. Our offsite storage was Google Drive but it no longer allows us to backup the rest of the lab.

Most of our data is coming from UCPath, ESR systems, and campus AP data. We are not true owners of our content so we are not responsible for backing up this information.

my biggest concern with using university systems is the university having control or access in any way to sensitive cultural data. The UC system does not have a good record with allowing community partners to have ready access to their own cultural data so I'm not sure I will ever store or backup my data to university-administered systems or equipment

Due to the high cost of backup software, most if not all of our backup solutions are open source and developed in-house with the exception of Crashplan for Desktop clients and AWS Glacier storage for DR.

Other tools, strategies, comments, or clarifications regarding responses to...

Subversion

The vast majority of our research data is backed up to tape that is archived off-site. We could recover essentially all of it in the even of a major attack (e.g. ransomware), but it would likely take months. Our tape system is near end of life, and we're investigating other options, ideally not things we have to manage within our department.

I use Time Machine to back up my laptop (including most of my research data) on external hard drives at both my home and my office.

Half of our data is stored on RAE dedicated resources (MyResearch shares) and I assume those are backed up by IT, but I'm really not sure. The other half (2-3 TB), are on Wynton, are not backed up, and we will be looking at options for backup.

We backup data from user home directories on our cluster. Not all data is meant to be backed up, some of it is temporary.

We need better back up options (e.g institutional crashplan subscriptions)

Github

We are using Veeam Backup and Replication

The most protected datasets in the Sali lab are P2 but almost all lab research is NIH- or NSF-funded with the expectation of it ultimately being made public. Consequently most datasets are only P1 protected and are duplicated in public databases and repositories such as GitHub, PDB, and Zenodo.

I'm a PI, I manage a cluster used by several labs, I have a semi-automated scheme for backing up 100 TB scale data to Google Cloud Platform

external hard drive

I am using Google Drive File Stream for basically all of my data. I am assuming that content is automatically backed up.

Our unit backs up its archives to a QNAP NAS which is backed up weekly to Backblaze. QNAP devices were targeted under the recent QLocker ransomware attack, which was not patched in time to prevent data loss.

My lab group has diverse levels of expertise and knowledge on this topic - from virtually none (especially in new members) to extensive. We need a single (or very few) solution and standardized training/workshops on best practices.

I am a big believer in multiple copies of data stored in different places. Thus ALL my researchers have a hard drive and are required to back-up there data there as well as on the Cloud.

I use Hash Backup for retention and backup. I use my own Gitlab server for version control of codes. The Gitlab server backups are handled by Hash Backup. All important information/codes are backed up to Backblaze. I have to pay out of my own pocket for the Backblaze backups.

Almost all of our data are on UCSF-maintained systems (e.g., Box, RAE, network servers), which I assume are backed up regularly.

We snapshot then incrementally backup hourly to a twin server (IGPP colo), daily to an offsite (SDSC) server and quarterly ("full" backupt) to AWS glacier.

We have a lot of different data streams. Some are not properly digitized (old video tapes). Many of the computers are backed up to one or more drives but not all of them are. Most are not also backed up to a cloud storage site which would be VERY much welcomed.

For backup of our local file servers we use Retrospect on a dedicated backup server and Windows backup for bare metal restore where necessary.

Other tools, strategies, comments, or clarifications regarding responses to...

Using duplicacy software for backup.

I find this survey very confusing. For example, when asked what percent of our data are stored on particular systems, we are asked not to exceed 100 %, but many of these data have images on multiple systems. The specific hardware details you've asked for are things that I, and I would imagine most PIs, have no direct knowledge of. While I can see why these data are valuable, I am concerned that this survey won't capture the information that you need.

Except for big genomic datasets, everyone in the lab uses their own system. It would be good to standardize this.

Our lab uses SyncBackPro to backup data to a shared Departmental Network Drive and local external hard drives. We also backup to Box, but to a lesser extent because it is a lot slower and more problematic. Individual researchers are responsible for their own backups with some doing them on a regular basis and others doing them irregularly.

Our tape backup system was developed in-house to meet our specific needs. It has been in use for several years and works well.

- We do some file level backups to USB and via rsync to a NAS box but the NAS box is not always online correctly (more a problem for server configuration than data). - Research data and daily observational data is copied to main SAN when final (daily for most). Daily rsync to NAS backs up ~80% of SAN but inconsistently due to space limitations and some NFS issues. - MAC laptops are managed by upper division and backed up by IGPP NetOps. - Windows laptops are managed directly by researchers who have USB drives for backups (unknown frequency) - Much of the older critical data is also backed up to USB drives but that is on an ad hoc basis.

I also use Time Machine and GitHub. GitHub contains code that can regenerate data tables that are not otherwise backed up by an outside party.

We just did a deep-dive into a single research department (Physiology) including their storage and back-up solutions. We interviewed a sample of labs and found storage solutions to consist of 1) Lots of external hard drives, 2) NAS, and 3) server(s) under the desk. "Backups" are done by 1) Manually uploading files that are under the file size limit to enterprise Box, 2) Backing up to external hard drives and possibly having a 2nd copy at the PI's home, and 3) Mirroring to NAS or another server.

We currently use an Enterprise backup solution know as "Quest Rapid Recovery" (https://www.quest.com/products/rapid-recovery/). It is not on the list of backup solutions in the questionnaire. System is on-premise with archiving to AZURE and AWS for immutability.

Off-site backup is to one of our sub-contractors.

All of these questions are about "unit-controlled" devices. My group is pretty dispersed, is not a "unit" in any sense of the word at UC, and so I'm unclear as to how to respond to this. I'm a Faculty PI, we have a lot of data for someone in the Humanities, and our backup is chaotic to say the least.

Our group server is RAID, which is a form of data redundancy.

My statistician does the backups and she works from home on DGSOM computers. She's supposed to do them once weekly but I'm not sure if she's actually doing them or what software she uses. As per our MOU with the company providing the data, we back up onto external drives and then store those in locked filing cabinets.

time machines

Data back-up is something we need to improve upon. We do not have a scheduled back-up time frame and we mostly use external drives and a lot of data doesn't get backed up. A new institution-wide back up plan would be of great benefit to us.

Most of our data is housed on drives / servers at the NBER, which does have backup mechanisms, but I'm unsure of specifics

Much of my data is stored offsite by a data management company Frontier Science Research Foundation. I also use SISTAT on campus to help with my research dat

Other tools, strategies, comments, or clarifications regarding responses to...

Interesting how when downloading documents using Safari, incomplete downloads are always in a "download" format which Box does not support. For that reason, there is a workflow hiccup where on my Mac in Safari, I cannot save downloads straight into my UCLA Health Box folders. Because of that, some files end up in my computer's "Downloads" folder waiting to be manually moved to Box filing, as opposed to allowing filing to occur right when I save.

A large portion of data is backed-up on external hard drives

Much of the work I do is public and is backed up by my collabators at other institutions.

I do not understand why I have never been able to EASILY get server space and easy backup for my data in my 21 years at ucla. It has always been so riven with red tape and limitations that it has not to be worth it. The onus is on individuals (or sometimes research units, if they are so organized). This is a burden on our time that takes away from productive work and it renders us vulnerable to data loss (which we have experienced).

Used to use Drobo, but kept running out of room on the drives and the drives sometimes got corrupted. Decided to use infinite UCLA Box storage instead.

I am not in charge of backing up

We have 100s of Tb of data. It's not efficient to back it up. We build in the capacity to rebuild the data bases if they are lost.

A big issue when I transitioned to store my data in the UCLA Health Box is that I cannot back it up with the Apple Time Machine. That was my way of backing up my data before...

Over nearly forty years at UCI, I have had a number of small datasets from surveys or human subject laboratory experiments, and co-authors also have their access to those small datasets. In some cases the entire dataset is included in final reports on paper. I have some original paper and pen survey results on paper. Later, studies were done online, such as on Amazon M Turk, and are contained in Excel sheets.

I started here in 2017 and was told that there were no servers to house data and should use Google Drive. No one has ever offered me backup options here. If I want it I have to pay for it myself because grants won't cover it.

We have deidentified data on multiple devices, so these could easily be restored if one of the devices is compromised. We are hamstrung on backup options by some old contract wording saying we cannot backup the data we receive to a cloud. Perhaps that can be renegotiated someday but for now we back up to an encrypted external hard drive that we keep physically locked up.

I am unable to provide meaningful responses to the questions asked of me in this survey, despite my wish to do so.

Most of our data are stored on either a lab server that is controlled by the David Geffen School of Medicine and backed through their services. Data on personal lab tops/computers are backed up/copied by lab members to this server. Typically NGS data are stored and backed up on a UCLA computing server, Hoffman2.

Data being backed up using external hard drive & stored in secure off-campus location.

We have our own NAS, maintained within our departmental server stack, which is backed up to my Google Drive account continuously, and is currently roughly 25TB

there is zero IT support at UCLA in the engineering school for faculty.

Recently bought on-campus backup NAS (Synology SA3400) for cold storage, but it's not really a backup (data are transferred to it to free up space elsewhere, it's not acting as a copy). We're still looking for options to add offsite backups + more consistent backing-up strategy (e.g. weekly, monthly) that isn't super expensive.

Back-ups occur irregularly, but at least on an annual basis.

Other tools, strategies, comments, or clarifications regarding responses to...

Whenever possible, data are deposited in the UCLA Library Dataverse at time of publication or other applicable stages for long-term preservation.

I recently joined UCLA and I'm still trying to figure out backup options

portable drives

Our lab uses a network server that is maintained by DGIT (David Geffen School of Medicine IT). We don't do the Backup. It is done by DGIT, and they don't share their information re: how/when they perform backup.

I am working on a UCLA Law server, on Dropbox, and on OneDrive. I am relying on our IT department and Dropbox to provide backups.

Here's my backup strategy: ~800 GB of data are housed on my desktop computer. This gets frequently backed up to an external hard drive using Time Machine. Occasionally, I copy all of this to the UCLA server as well. This is not an automatic thing, so it is labor-intensive and time-consuming. If I could set that as a Time Machine drive, my life would be so much easier. An additional 2 TB is housed on an external hard drive because it won't fit on my desktop. That is frequently backed up to my external hard drive and the UCLA server. This is a manual process that is time-consuming and labor-intensive. Having an easier way to do this would be wonderful. Whatever options you consider, PLEASE make sure that it works automatically and smoothly with a Mac. It would be preferable for it to serve as an off-site Time Machine backup drive somehow to utilize that built-in system.

We use g.ucla.edu for 1-2Tb of data and it is backed up real-time. All files are on Google Drive.

What exactly do you mean by "unit" e.g. unit-controlled NAS? We back up manually or semi-manually to a departmental server. We also backup locally to external encrypted drives

I'm a mac person. so I use time machine to back up my data

Box is cumbersome to use when collaborating with non-UC colleagues, which is why I maintain a personal dropbox subscription.

It is very hard to beat Wasabi. No ingress or egress charges. Excellent customer support. Its a good deal.

We've recently started to get large amounts of genomic data. We bought a second server to house original data, and do a bit of backing up there. But it has limits. So we're backing up so little because we can't afford, and the university does not provide, back up, even for the new Borg storage, which we have to pay into at a pretty hefty rate. That seems inexcusable to me.

I have thought of using a cloud based service to abandon my lab's server for the sole purpose of having our data backed up in the cloud. The downside for me is that of privacy/hacking with such cloud based applications. So what should we do? A UC-wide server/or backup cloud option?

Old-school "sneakernet" backup. Data are physically transferred from one machine to another via USB drives (usually) or from desktop or laptop to USB drive.

This includes only what we backup for users, but we tell them too make their own backups, as well as store their primary data.

Retrospect

I backup my personal computers between them and between an external drive. All the research, instructional and administrative data, which I have uploaded on the cloud, is also backed up on my computers.

Every 5 minutes FTP transfers from instruments to storage server. Storage server gets uploaded to AWS Deep Glacier on an hourly cron job via rclone.

We use amandabackup to backup nightly updates and the full system once a week.

Other tools, strategies, comments, or clarifications regarding responses to...

Whenever possible, we publish our datasets alongside our papers (e.g., using the Open Science Framework or Dryad)

We are currently using the College of Engineering Bacula Backup. This is only useful for servers and statics desktops, however it is not compatible with all OS's, for example new OSX releases like Big Sur and above. "Off site" would be a tape backup to a different building 200 yards away.

We run several Ceph installations and also several Eucalyptus private clouds. The Ceph installations here at UCSB are all using 3-way replication. Our backups cross these installations, either directly accessing Ceph or using Eucalyptus storage that is hosted in Ceph.

Backups are on external USB storage; all data is stored on both Box and Google Drive for multiple redundancy.

Data is also stored on UC Qualrics account

None

I am backing up to a personal external harddrive for my laptop and to Google Drive for files on the desktop. I plan to get an external harddrive for the desktop in the next week or two.

I also back up to a portable hard drive on-site. This isn't mentioned among the options for backup.

I have not backed up older data that resides on redundant hard drives. This is a problem, but I have hesitated to place everything in Box and Google as I read that the services will likely reduce quota for our accounts and I need space for new data.

our NAS was configured by our IT services to automatically backup to offsite/cloud but I don't know details. however our users have to push their data to the NAS for backup

We have really like 3 classes of data. Rawdata is the imaging from the microscope. Processed data is the results from operations on the rawdata. Other data which would be something like a movie or graphic that is manually produced from either the rawdata or processed data. In rawdata terms we have several hundred TB of rawdata like between 1/4 and 1/2 all data, Processed data is the bulk of it and Other data is likely in the hundreds of GB to single digit TB range. The raw data can never be replaced, processed data can usually be reproduced from the rawdata and other data is usually pretty unique in how it's produced so can likely not be replicated exactly if at all. We used to keep off site nightly snapshots but the cost of storage exceeded our $$ allocation so we resorted to nightly local snapshots and highly redundant pools with multiple hot spares. and dumping all our rawdata on google drive. As to the ransomware question there are some uncommon scenarios that could possibly encrypt a pool but the attacker would need direct access to the filesystem (i.e. not over NFS or Samba) and would need to have an understanding of how ZFS works. Not saying it's not possible just that it's not your usual scenario. VM's and servers are provisioned and maintained in code. Any unique data for that system is kept in a mount on a zfs filesystem. We try to do it in such a way that you can blow away the system and be back up and running in about 30 minutes.

submitting data to NIH mandated data repositories

On-prem we use Veeam Backup & Replication. Disk-to-Disk-to-AWS Cloud backup.

We're a new lab and I would love to have a backup strategy but it's a lot of investment and thought to figure out the details and put the plan in place. Until data gets lost it's unlikely that we're going to have a good solution

we use retrospect

I believe Google Drive storage includes backups, but I'm not 100% sure of that, and unclear whether you're asking only about independent backup systems here.

The backup information on my NAS has redundancy from a RAID 1 and RAID 5 configuration, but they are connected to the UCI network and in the same room as my working desktop computers - an obvious vulnerability. The network speed in my building is only 100Mb/sec so transferring vary large amounts of data to an external website would be very slow.

Other tools, strategies, comments, or clarifications regarding responses to...

Using Bacula Community Edition to back up to on-campus NAS devices. There's some attempt to back up systems to NAS's that are not in the same building, but no true off-site backups.

Long-term backups go not to tape but to bare drives in a fireproof safe

We have tape drives we are in the process of implementing but it will be a while Typically we are using Veeam R2R gets a copy of all our research data after each cruise.

we have hard copies of lab-internal data we have microsoft excel files backed up on multiple computers

MY data consists largely of manuscripts, talks, mailing lists,and correspondance.

I store my data in at least 2 different computers, often on different computer types (linux vs PC). I copy files to network drives that are backed up nightly by UCI Health. But that doesn't happen regularly. I used to back up regularly onto external drives, but changes in hardware and lack of computer support to program the backups ended that.

As I said, I work independently of any "research unit." I store my data on my laptop and one a "seagate backup plus." I also print out almost all of the data that I generate.

I was required to have data backup plans for NSF and NIH over ten years ago, long before UC provided any backup infrastructure or knowledge. Individual researchers were left to come up with their own ad hoc solutions, and UC is very late to the conversation.

RAID 1 with ZFS

USB drive.

I do quarterly back up myself, to an external hard drive, with selected files also backed up to thumb drives

Also back-up with external hard drives

I and many of my colleagues routinely make research data publicly available via repositories like GitHub or Open Science Framework

What we have as a backup system is basically a set of rsync scripts that I wrote many years ago that create mirrors of data on multiple drives on the same machine, and on different machines, including off-site machines. They run as cronjobs nightly, weekly and monthly.

our "offsite" backups are at SIO, so they are not *far* offsite

i have no idea what a lot of these things are

Everything important is on desktops multiple desktops and laptops locally as well as either Dropbox (mostly) or Google Drive (slowly moving that way). I think if Dropbox/Google were hacked I could get almost everything important from the local stores. Not sure if that counts as "backing up". Probably not. I rely on the cloud services doing back ups.

Most of our data us backed up by the UCI Data Center and I am not familiar with their backup systems.

NetApp appliance, Hadoop cluster. Not really enough space to back things up, but netapp is used to save and store a fair amount off research servers (and host home directories for this unit). Hadoop cluster is too large to be backed up and would be 100% lost (and has been, several years ago; the cluster is very highly protected now).

We also use VEEAM. Daily incremental backups, Weekly Full.

Other tools, strategies, comments, or clarifications regarding responses to...

Unit on-premises storage is primarily built on top of ZFS storage arrays with daily snapshot backups to both on-premises and recently to cloud platforms.

I will clarify our backup strategy as I don't think the questions truly reflect our process. We have research instruments and data workstations that need backup. These are backed up via cloudberry to on-site DataRoboitcs Drobo devices. These data are also then backed up to AWS into glacier storage. Onsite backups are retained for up to 2 years depending on storage limitations. Offsite storage with AWS is kept indefinitely.

na

na

NA

Practices vary greatly depending on the researcher at the law school, and the specific research project data requirements.

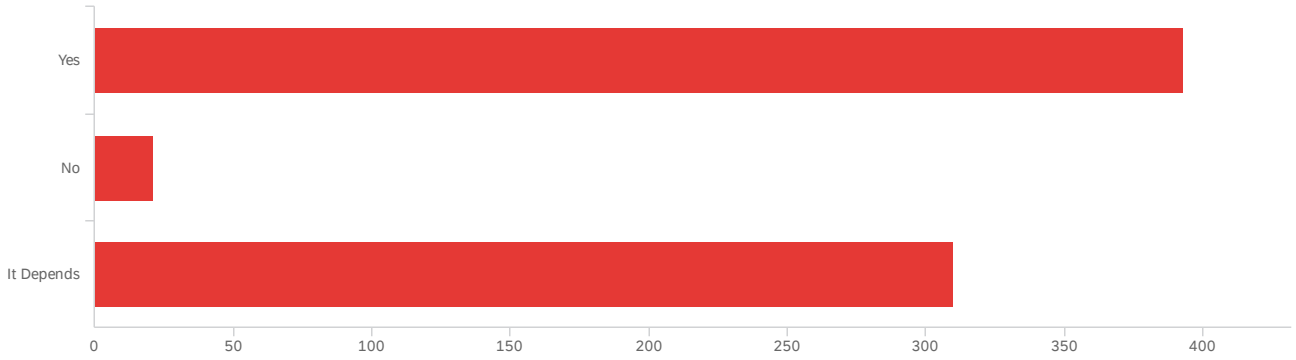
We are running on prem Crashplan server (backup select remote laptops), on prem IBM Spectrum Protect (backend/Unit IT file level backup), Microsoft Shadow copy (user facing backup), Quest vRangerPro (VM level backup)

Most research data within Engineering are managed and maintained by each individual PIs. We may have a small amount of research data that may be stored on our servers and that may be web pages that refer back to their research and data they may upload to their site.

Also doing snapshots (hourly, 6-hourly, daily, weekly, monthly) on central NAS device. Additionally, Digital Preservation (long term archival, multi-repository storage of core data and metadata) in use.

ma

Q4.8 - In your unit, if you were offered a centrally funded or partially subsidized research data backup solution, would you use it?

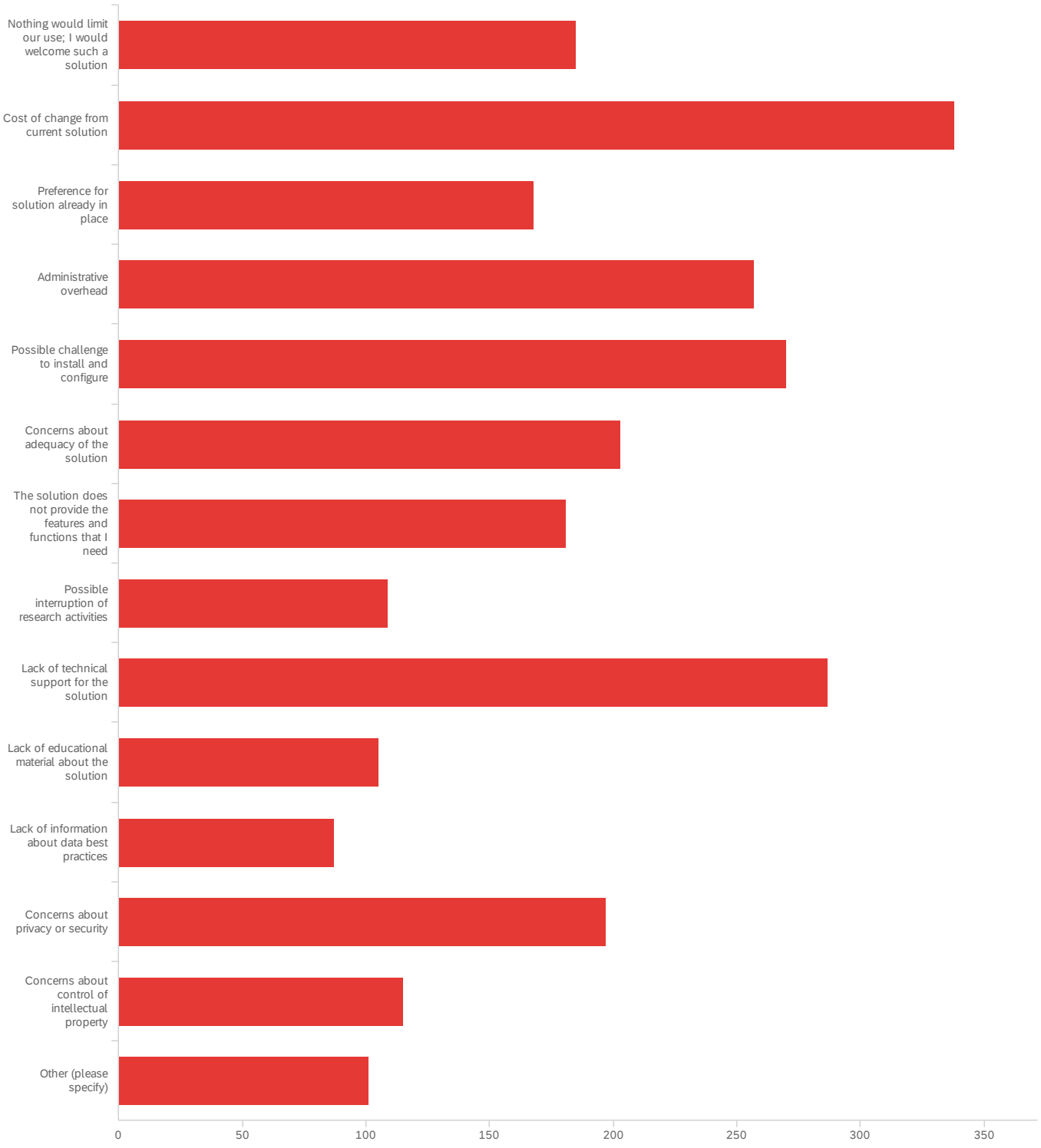


#	Field	Minimum	Maximum	Mean	Std Deviation	Variance	Count
1	In your unit, if you were offered a centrally funded or partially subsidized research data backup solution, would you use it?	1.00	3.00	1.89	0.98	0.96	724

#	Field	Choice Count
1	Yes	54.28% 393
2	No	2.90% 21
3	It Depends	42.82% 310
		724

Showing rows 1 - 4 of 4

Q4.9 - In your unit, what might limit use of a centrally funded or partially subsidized research data backup solution? (Select all that apply)



#	Field	Choice Count

#	Field	Choice Count
1	Nothing would limit our use; I would welcome such a solution	7.11% 185
2	Cost of change from current solution	12.99% 338
3	Preference for solution already in place	6.45% 168
4	Administrative overhead	9.87% 257
5	Possible challenge to install and configure	10.37% 270
6	Concerns about adequacy of the solution	7.80% 203
7	The solution does not provide the features and functions that I need	6.95% 181
8	Possible interruption of research activities	4.19% 109
9	Lack of technical support for the solution	11.03% 287
10	Lack of educational material about the solution	4.03% 105
11	Lack of information about data best practices	3.34% 87
12	Concerns about privacy or security	7.57% 197
13	Concerns about control of intellectual property	4.42% 115
14	Other (please specify)	3.88% 101
		2603

Showing rows 1 - 15 of 15

Q4.9_14_TEXT - Other (please specify)

Other (please specify)
Hard to answer this question without knowing the proposed solution.
My PI is super cheap. Really it would just come down to cost per TB.
UCSC would outsource this to individual faculty without any kind of information or support.
Increase cost
Whether or not it supports receiving ZFS snapshots
Needs to be flexible
The key issue for us is whether it would meet with the data security requirements of our many data contributors (30+ data use agreements).
If partial subsidy meant it was still expensive.
Difficulty in implementing across teams and existing data management systems used by the unit.

Other (please specify)

I think it unlikely that a university-wide backup solution would not meet security/technical support needs, but noted just to be complete. We would be EXTREMELY excited to participate.

We would likely use the campus setup once it is mature, while keeping our current solution (just bought new hardware this year). If the campus setup worked well, we probably wouldn't buy new hardware when it fails.

Concerns about longevity/stability.

durability of the service. I'm tired of cycling through multiple expensive, difficult to use campus solutions that change from time to time.

I don't need such a service.

Cost

cost if only partially funded

future ownership of data / code

Google/cloud has the best and most secure and most convenient technology

UCSC campus is subject to frequent blackouts

Lack of staff help/funding for staff

Whatever solution cannot be Windows centric. Should support macOS and Linux equally well.

I built a good solution myself

Regulatory compliance of solution

I would certainly welcome such a solution, but many of the concerns listed require careful thought or investment for it to work well.

Asian language support

Would be a college-level decision

Ease of access to data for running analyses

Ease of sharing data with collaborators outside of UCLA

why isn't RAE sufficient?

Ability to use from Linux systems using standard tools (rsync, rclone, sshfs, etc.)

Data transfer and access speed. This would be a HUGE concer for large image datasets. 10s to 1000s of GBs.

Multiplatform support

Other (please specify)

To be honest, I find this entire discussion very amusing. We are hurting all cross campus by not having storage for even the first copy of our research data, leave one a second copy for backup. In my mind, this discussion is divorced from reality.

concerns about: "You agree to indemnify UC San Diego against any third party claim based on an allegation that UC San Diego has violated that third party's copyrights, trademarks, other intellectual property, privacy, publicity or other legal rights due to your actions or omissions."

Very Slow Internet at the current office/field station.

Dealing with unmotivated and often unhelpful UC IT support

If it doesn't work seemlessly from the command line on my Linux machine it is not useful to me.

Concerns of being able to easily access backups as required.

This is a HUGE need for campus. We are a core facility with expertise in big data and systems administration, but all the labs we work with (>50/year) need a centralized, low cost, secure and backed up data storage solution.

Anachronistic archaeological workflows

Our backup solution is tightly integrated into AWS which would be expensive to detach and move to another backup solution.

We've already budgeted data backup in our grants; why pay double?

It all depends! It would need to be well supported on my platforms and work well. I can't summarize the criteria easily using your checkboxes.

The current offering only provides 1 month of backup

cost

concerns about limited permissions/access

some computers are not on the web

Used university backup services in the past and when I needed to restore data was told they could not restore my data. After paying for the service I could not get my data. With costs at ~$50/TB I see no reason to trust anyone else with my data.

Integration with compute cluster

cost

user friendly

Cost of the solution itself

permission from data provider

I already use Crashplan, provided by UCSF

Other (please specify)

It would depend on what the solution is. We run solaris 11, often that's a problem. Since we use ZFS, we would probably use your service instead of glacier (which is already cheap and easy to use)

If the proposed alternative doesn't address these concerns, any of them could be troublesome, but without a proposal, I really cannot answer the question.

Speed of access, UCSF network is too slow

One issue is speed of access and download. Is it simply archive with relatively slow access or can it be used for data that is being actively analyzed.

As the sysadmin, I would welcome a more secure solution but don't know what the research funding will support.

DUA agreements in place do not allow for any networked backup services of PHI/proprietary data.

It would be impossible to change our MOU at this point, but we could probably switch over for future projects if the external entity providing our research data will agree.

HIPAA secure

I don't know enough to answer this question.

Download and upload speed

Compatibility with my platform (bleeding-edge Linux, often not supported)

Constraints on allowability of research data backup solution for specific datasets (HIPAA-protected, etc.)

our institution previously offered data back up services that we relied on. I turned out that it was not being actually supported to the level advertised, so when we needed to recover our data, it was not available to the extent we were led to believe. We resolved not to rely on university guarantees for mission critical services, but to buy our own hardware and hire our own staff to have control and to know exactly what was being done.

price to pay by my lab

Meeting required guidelines for information security

I am retiring from UCI

contractual restrictions on where we can store data

future cost uncertainties

Most of our data are secured by IT services or the core that manages the computing server, so not sure how another service would add to their safety measures.

speed: our NAS is very fast within the building internet

cost to use (not cost to change)

ease of access from different locations

Other (please specify)

Concerns about sensitive data

CPU overhead in syncing

g.ucla.edu / Google Drive is useful for almost all of our needs. It is easy, "free", and unlimited for modest data storage (1-10Tb).

Ability to access remotely

inability to access when local connections disrupted

cost

This solution would likely be prone to hacking, just like all other centralized backups.

If the solution does not support Linux, Mac, and Windows they it is useless, because that is what the data is stored on. Similarly if it can't support roaming users like on laptops they it is worthless.

Ease with which the data could be transferred to the backup

Sufficient configurability for computer science researchers

HIPAA compliance, for some of our data, and difficulty/ease of use

OS support. We need readily available, reliable clients for macOS and Linux. My experience with proprietary backup solutions on Linux has not always been good.

I would need help moving all of our existing data. I cannot currently afford to switch to unit-driven solutions. It would be a challenge to have different groups working with different solutions. We all got used to working together in the google suite.

Diversion of my time to deal with a new entity.

Applicability to nonstandard systems

Limited network speed to backup 20 terabytes of data.

bandwidth to transfer data to backup

Don't know

Cost; we would not be willing to pay a monthly fee for these services.

I don't know

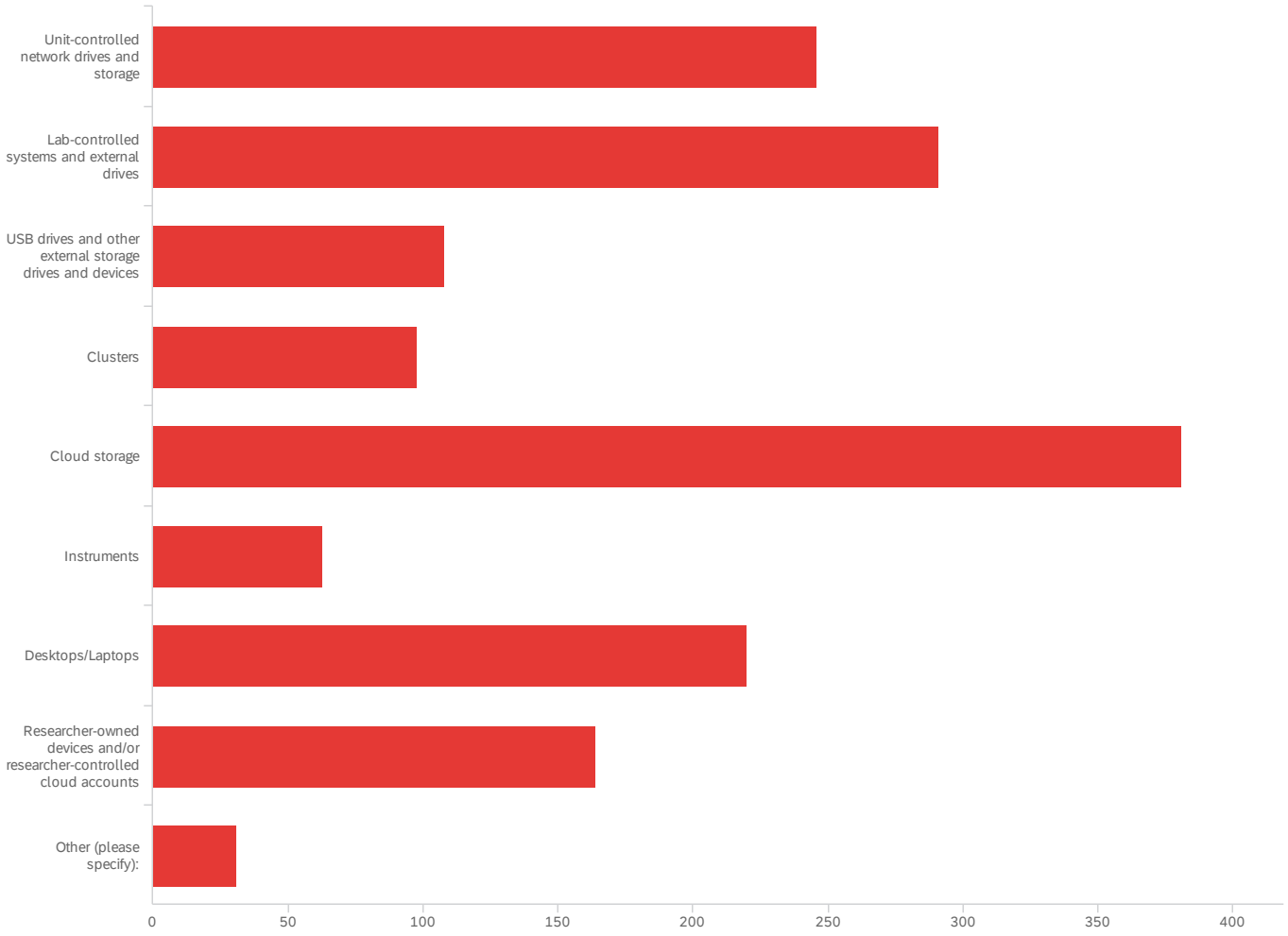
ease of use

The biggest challenge would be to transform our current backup method. As well as transfer of 100Tb of data from AWS.

Concerns would really depend on the details of the centrally subsidized solution!

Other (please specify)
Ability to restore data rapidly

Q4.10 - What situations do you think should be prioritized for a system-wide research data backup solution? (Select all that apply)



#	Field	Choice Count
1	Unit-controlled network drives and storage	15.36% 246
2	Lab-controlled systems and external drives	18.16% 291
3	USB drives and other external storage drives and devices	6.74% 108
4	Clusters	6.12% 98
5	Cloud storage	23.78% 381
6	Instruments	3.93% 63
7	Desktops/Laptops	13.73% 220
8	Researcher-owned devices and/or researcher-controlled cloud accounts	10.24% 164

#	Field		Choice Count
9	Other (please specify):	1.94%	31
			1602

Showing rows 1 - 10 of 10

Q4.10_9_TEXT - Other (please specify):

Other (please specify):
Some people still use some USB external hard drives for this - that is not a good solution and should be the firts to be migrated to a central, secure system.
Please, don't make us select stupid choices: if you're gonna do something, do something for ALL OPTIONS
Do not know
A storage is not a backup solution. In order for backups to work it would need to be automated and self-serviced
All data
fast switches, this will be rate limiting, can't use cloud if a slow switch is a bottleneck.
Campus provide storage for the first copy first before wondering about backup.
Speaking a bit for others in my department. Few run clusters and network storage. Many are scholars with books in progress on their laptops. You would be surprised how vulnerable some of my colleagues are.
Not sure
Core Facilities where the end user may not be responsible
OS/desktop image backup
Researchers need the freedom to identify this, a one-size-fits-all prioritization will probably fail
I'm not sure the technical name, but a remote file storage system operated and maintained by campus that is secure and that I can ssh into to access all my data (ie. what Fred Hutch Cancer Research Center has)
by value of material not by storage method
Backup of extremely large datasets
RAE
Unit devloped applications and systems
Not sure

Other (please specify):

I don't know enough to answer this question.

It really depends on the data, it's use and priority, and the confidence we have in the system. Frankly, I don't trust any campus administered system.

access speed (bandwidth)

don't know for sure what is meant by unit

A way to use Time Machine with an off-site centralized backup.

I don't understand this - these are devices, not situations

Servers. a removable drive is by nature impossible to backup if it is removed.

Something that backs up by desktop automatically.

Research data that is not currently backed up in a secure fashion.

Linux servers

on-site storage is risky due to risk of earthquakes, we need both instrument/computer storage onsite, offsite and on the cloud. I need to be able to access my data from various locations.

I have no idea

I'll leave answering this to the experts

Q4.11 - What future plans do you have for backing up the data in systems under your control?

What future plans do you have for backing up the data in systems under your...
Google Drive is going away, we currently back more than 500T to GDrive.
Plan to back up every few months
Migrating from Box, which the university is abandoning, to Dropbox, and probably using iCloud backup of external drives synched with the Dropbox. But this is a serious problem for us.
hard drive stored in another location
None
We are considering working with National Data Centers for a solution for Lick Observatory but would welcome a UC solution.
Purchasing of external hard drives.
Our current system is adequate for the foreseeable future.
Continue to use external back-up drives
Learning more about how our unit data storage fits with the Center's research storage and the Institute's research storage, and what their future plans for data storage are, including use of a centrally funded or partially subsidized backup solution.
I have thought about requiring manual input for my 3rd stage backup. This would completely eliminate the possibility of my data being unwittingly compromised. But requiring manual intervention carries its own risks and would require personnel redundancy.
I need to know if the AD server is being backed up. If not, we will have to find a way to backup our data.
We want to get server space and/or cloud storage space but we are trying to find out how we can get that without spending the money on it, because we do not have the money, but on the other hand the data we are generating is essential to multiple research projects
Need to discuss with IT
Offsite backups and replication. Ideally we would reintroduce tape as a component of our backup system (decommissioned due to cost).
We will continue with our present system
We back up to disk and then to tape. We'd like off site storage other than my home fire safe.
Move data to long term archival storage after i graduate
Whatever is available.

What future plans do you have for backing up the data in systems under your...

continue as is

Our current solution seems to work and are not planning any changes. The only drawback to having a unit-controlled system is if I need to do something (e.g. restore a deleted file) I need to work through them and do not have the ability to do it myself.

Continuing current practice. In addition to backing up via networked storage, I keep copies on multiple external drives that are kept offline.

No plans

Continue to back things up manually to external drives and the cloud

Our backup strategy economically and efficiently guards against data loss to a reasonable level of assurity.

Once we have migrated our file server to Wynton we will setup daily backups on an external server (our current file server). Long-term, Wynton say they will offer backup solutions.

Move unit server management to central campus management.

Expand our tape system to backup to Amazon Glacier, to offer researchers dual live copy + local tape backup + cloud-based backup.

continue as is

All research data are saved on cloud services (Dropbox and Google Drive). I might back up DropBox stored data with an external drive.

continue to use Crash Plan / Code 42

We are in the process of developing an offsite backup solution at UCB that backs up our data currently stored at UCLA. Required us buying a big NAS and colocating it at UCB.

None - waiting to see what happens.

I am waiting for guidance from the university. No plans to change.

Shift more to cloud storage if/when available from institution.

continue with Scripps IT

In the absence of a campus wide system we will use cloud storage (Google Team Drive) for all our research files until we hit maximum storage. Then we will have to move to local RAIDed hard drives or S3/Glacier backed up hard drives.

Same as current

We will continue to use current solutions unless another one presents itself. We have not chosen to purchase backup directly through SDSC because it is not cost effective at our level of data (i.e., we do not have enough data to benefit from economy of scale).

We continue to evolve and do what we can to provide solutions appropriately tuned to our researchers' needs for the best value we can engineer.

What future plans do you have for backing up the data in systems under your...

really hoping google drive stays, that would cover the vast majority of our needs. We also may be generating large data microscopy files - but instrumentation data is easier to store on an external. What I worry about is managing all the student's / staff data. Individuals can't be trusted to reliably back up their data and I've seen a lot of horror stories. Solution needs to be easy to enforce for all students (i.e. all data goes on google drive no matter what) otherwise it won't be sustainable.

Personally financed cloud storage using CrashPlan and pCloud.

continue current system

I'm going to keep muddling along until a functional cloud-based system is developed that I can use. I'm already being asked to cover costs I never thought I would be responsible for when I came to Berkeley (like Grad student stipend enhancements). The cost of a back-up system is important to me. When I have funds, I can purchase multi-TB backup systems for my office, but I'm not interested in assuming an expensive cloud-based service with annual charges unless they are modest enough that I don't have to worry about losing my backup privileges if I don't have a grant.

Nothing concrete in place yet, but there is a need to back-up existing database to the cloud

hopefully, external hard drives and cloud back up

None. External hard drives are adequate.

Purchasing cloud subscription

We offer backup services for research data as part of our DataONE research data federation to a national community of researchers.

for the time being, we use 2 Tape Libraries (LTO-7 tapes) that ECI manages to backup servers and VMs. Bacula is the software used for our backups to tapes.

I would like to have off-site backups on systems that do not require configuration/administration

NAS

Backing up all physical data remaining from months prior to remote work onset to department-controlled network drive.

Would like triple redunadancy instead of double.

Expanding our backup capacity and capability.

Same as usual

We are looking to UC to help with a university wide solution to this important issue

I do what IT recommends I do as far as backup.

N/A

external drive

I don’t have any

continue to use personal NAS

What future plans do you have for backing up the data in systems under your...
none
none at this time
ask me next year
Off-site an additional synology NAS to have geographic redundancy.
Moving all backups onto google drive.
Google Drive, essentially.
Not sure yet.
No change planned.
acquire a lab-owned storage cluster
None
Hardware refresh.
continuing to backup the same way as during the last 12 years
I personally stopped storing data on the PC desktop and use K-drive that is backed up daily. I encourage my lab members to do the same. However, I also have Linux workstation that houses data that I do not know how to backup in centralized way. I use an external drive to back it up irregularly. As far as I know, health campus does not support Linux computers at all.
n/a
personal hard drive or cloud storage
Using UCBackup more while phasing out Veeam
We don't have plans to change our current approach.
I would prefer a cloud based solution.
I would like to find ways to migrate from lots of external discs (totalling nearly 1 PB of data) managed by individuals in my lab to a combination of cloud based and local centrally managed network storage.
planning to speak with our department specific ITS people to put something better in place than we currently are using
N/A
All our important data (microscopy, sequencing, articles related data) will be on a NAS and a server that will both be backed up using AWS (most likely). Individuals use their own backup systems (hard drives, google drive, box, dropbox).
None

What future plans do you have for backing up the data in systems under your...

I use Github under an academic license, google drive for temporary documents, and dropbox for course materials.

None in particular

I tried to migrate all my data to one drive but found it was way too awkward compared to Dropbox re access and file management. So now have everything split with some on one drive some on Dropbox

Continue keeping much data on OneDrive and desktop data backed up to Carbonite

None unless recommended otherwise by college's IT professional

no plans, maintain my ad hoc systems

Continue to use AWS, crashplan and github

Advocate for the use of the UC Backup (Commvault) solution.

For now we are using only Sharepoint for backups, no future plans currently.

None. I have a solution and hopefully I can continue to pay for it. It would be great if UC could provide more support for data storage and back up instead of leaving us to fend for ourselves and create our own systems.

We do work with the ITCS Data Center. However, this is expensive, but because of data security concerns we are still with what campus offers us.

We are currently only using systems that are backed up by the University.

Not sure

Same as now: personally purchased portable disks. The backup options provided at UC Merced don't support Linux systems in a convenient way.

If I can get resources local servers would be a first step, but these will still be of limited utility since our switches are slow 1Gb connections speeds in many of our research buildings.

Continue same backup solution; I needed to create it since there wasn't support for it for our research group.

I assumed I would gradually move all data storage (some of which I still have in Dropbox) to Google Drive where we (UC) currently have unlimited free storage. I assumed this acted as a sufficient "backup". I would also consider using SIO IT/CoLocation Server Facility network storage.

Copy all data to box

We can barely afford the storage for 1 copy. So we will not have backup as the decision is between doing new research or backing up research. As argued before, the entire discussion seems eerily divorced from reality.

We backup what we need on our own.

investigate further off-site backup solutions

I have been looking for a centralized institutional solution that will live on (Library, etc.) for ecological data so this survey gives me hope.

What future plans do you have for backing up the data in systems under your...

Use Cloud-native tool for data in the Cloud. Use a SaaS solution for everything on-prem and potentially for the Cloud where cloud-native backup is not granular enough.

None. My cluster data sets can be reproduced if needed. Network servers/rsync back up primary research data, and Google drive (using duplicty on Ubuntu) takes care of my notebook.

N/A

Same as outlined here, no change

Presently, in involved projects, data are managed by the principal researcher. I will be principal investigator in 2 projects. In both, data will be backed up in cloud services, such as AWS

I will continue using a combination of Box (for HIPAA compliance) and Dropbox (through another university account) until a better solution is provided.

Buy more hard storage

Without university support, we will install a local hard drive backup plus some periodic offsite backup (which may be as simple as manual backup to storage that is taken offsite).

Continue with Druva if it is available

Continue with more external USB drives

no new innovations

I am just starting my research lab, and I am very eager for a unit-wide solution for backing up data

None

We do not have any changes in place, but would likely welcome a system wide research data backup solution as handling our daily backups for a year is clunky.

We will continue to use AWS S3 and Glacier with versioning until a system wide solution is in place. Would be great to be able to get a discount on AWS storage as part of the UC system.

A campus licensed solution would be welcome

None.

Continue triplicate storage: External HDDs (not connected normally to reduce risk of ransomware attack), cloud (google), and local computer(s) which are used to also analyze data

We are happy with the systems we have. Future plans include replacing aging hardware and updating software systems.

No changes.

Nothing beyond Box at the moment, but we are trying to understand the "jungle" out there in terms of real time access to secure storage as we use the data almost daily.

What future plans do you have for backing up the data in systems under your...
backup more frequently. With IT support, we hope to set up automatic backup.
I need an off-site backup solution for our NAS.
i have no heard any plans for future back up.
Have external drives in the office and at home so data is stored in multiple locations.
Continue use of Google Drive
1. Figure out how to backup Google docs/sheets/forms etc from the Google-drive to the external HDs using rsync. 2. Reinstate daily backups...
Get our own server functional and set up automatic backup in the cloud
use system-wide backup solution once available
using current solution
using current solution
Using external hard drives
AWS
Using CRSP
I will continue daily backing up of new data onto the lab's Google Drive.
External hard drives
More low-cost cloud-based backup solutions.
No planned changes
The current arrangement is working well for us so far.
Plans currently in flux, exploring central IT offerings.
Continue to use a 4 Tetrabyte Hard drive at home to back-up my laptop and continue to use the campus CommVault system to back-up by desktop workstation (iMac) in my campus office.
Our most recent ideas are to create a home grown script that would backup our Wynton data to our RAE Shares. This backup would not be encrypted, and the success of it (from a Ransomware protect) would depend on whether RAE is being backed up to encrypted servers not accessible to intruders.
Backing up regularly and more frequently.
I plan to continue backing up my data on Dropbox and on external hard drives.

What future plans do you have for backing up the data in systems under your...
Designing DR system for cluster storage now
More of the same
Working to create a database for our data based in google drive or box because they are the easiest to interface with
We're looking at Veeam's new immutable backup option and will evaluate that against off-line storage on hard disks.
If no central solution is provided at a similar price point to our self-managed backups, we will expand our existing offsite backups (which currently only cover our A3 and most A2 datasets) to cover all A2 and A1 datasets. (We are reluctant to spend money on this at present due to the uncertainty around a central solution being provided or possibly even mandated.)
We're currently moving our lab's combined data from (unlimited) box to lab-controlled NAS drive (primary copy) and AWS (backup, off-site)
If Box limits are put in place, it would be difficult to then download all the material (mostly large images) that I uploaded to Box, on the advice of the UC-Berkeley IT people.
More automated incremental backups of local data sets to GCP
Keep the same way to backup the data as we've done
Our lab is heavily dependent on Google Drive and our SPA account. We produce tons of imaging and genomics data. And having all of it in the same place, easily accessible to everyone is quite key. I have no plans at the moment, as I am only learning NOW that the unlimited space is not going to be an option anymore
Cloud storage + external harddrive
I recently started my lab and am still trying to find the best solution for backing up data, hoping we can coordinate as a lab in the next few months
Local solutions for now but would prefer low cost centralized, offsite
If an economical enterprise backup solution becomes available we would happily transition our archives to centrally managed, redundant storage.
No special plan than keeping using the current structure
I was about to purchase a backup system
none
None -- more of the same
Savio IST storage: https://technology.berkeley.edu/storage
NAS with sharepoint backup
Closer consultation with UCLA DGIT on ease and speed of data access
No change

What future plans do you have for backing up the data in systems under your...

No plans

Would like to get cloud but currently prohibitively expensive. Waiting for UCLA to provide affordable options. I believe they are looking at AWS.

I do my own back ups for my own datasets.

honestly, I back up my Dropbox onto an external hard drive now and then, plus have some other data on the external drive, and I rely on Box's and Google Drive's own backups. Just haven't had the time to revisit this and see whether there's a better way....

UCLA Anderson IT says they are implementing a solution for faculty, but I don't know the details yet.

None at the moment, although I would welcome system-wide improvements. That is actually something that I wanted for many years. I finally have a backup situation that is pretty good. The software (SyncBackPro) is highly customizable and does what I want it to do. It is very easy to backup to a Departmental network drive (not exactly off-site but I'm pretty sure it is not in the same building as our lab). I would like an easy way to mirror those backups from the Network Drive to the cloud to have truly off-site backups. I have been able to do that using Box, but it is generally quite slow and sometimes has problems.

We are working on a transition of at least some of our operations, including the main data storage, to a cloud provider. We currently plan to keep a complete copy of our data locally and keep it up to date with rsync (although this is subject to change, depending on bandwidth transfer time and cost).

Disk to disk backups with offline drive sets (zpools) in rotation (bi-annually).

I will continue to use GitHub, GDrive, but will likely also use hard drives for sensitive data

Duplicated offsite backup

Reliance on vendor-provided solution that is overseen by ITS

adding portable copy and cloud copy

We intend to continue with our current strategy

we're probably going to continue as we currently do until there is a coherent, cost-free plan for research data backup. Since there's effectively $0 in a humanities research budget, making this backup cost-free, at least for those of us without NSF/NIH type grants, would be unbelievably welcome. Even more welcome would be a coherent, easy-to-use interface, both for backup and restoring lost/corrupted data.

Expanded use of on-site NAS RAID array storage

My student has proposed setting up a lab server for data repository

Google Drive or another cloud-based resource

Planned migration from current ITS managed monthly backups to alternative UC Campus datacenter to a fully AWS cloud storage solution.

No change from current protocol, as we are limited by DUA stipulations to back up only to non-networked media/computers.

Trying to back up more frequently to the Cloud

I will continue to use cloud services and external hard drives.

What future plans do you have for backing up the data in systems under your...
now that I've completed this survey, I'm going to double-check our practices
Nothing different from what we're doing now.
External drives
Saving all standard computer backups to cloud storage and automating a system to backup research data to cloud.
I would like to be more organized - right now it is a patchwork of Dropbox, Google Drive, etc.
Public use repositories
Having data on two independent backup drives
I need to think more about backing up our lab data. Having guidelines and sharing of best practice (e.g., info session, webinars) would be useful.
I print stuff off. Put hard copy of book/article I'm working on in file cabinet. Sometimes I email myself work-in-progress.
Continue keeping all documents digital and saved in our departmental network drive (UCLA Health IT maintained and backed up) and UCLA Health Sciences Box folders.
none
Same plan. External drives in addition to what is on computers and the cloud.
Backing up on secure external harddrives
Use the kobox
None at this time.
I will continue with the current solutions: Box and Carbonite Customer Support.
I will continue to use non UCLA cloud services since they are cheaper and less bureaucratic
Researchers back up their data to Box and use either 1 to 2 external drives. Wider implementation of automatic backups to external drives.
Get grant funding, buy big hard drives.
Continued off-site hard drive backup, possible move to cloud.
Not under my control
We have code bases and access to raw data needed to reconstruct much of our data base in the event of loss. We back up short term mission-critical data and historical data that cannot be easily recovered. We have minimal capacity for offsite backups, except to duplicate data onto equipment of collaborators at other institutions.
Do manual back-up to an external hard-drive. But I usually forget to do it regularly.

What future plans do you have for backing up the data in systems under your...

Purchase additional storage servers

n/a

We currently have everything backed up utilizing whatever system DGIT has in place.

I'm going to back up everything on encrypted hard-drives

None unless it is provided by the institution. Except for prayers.

nothing specific at the moment

On the basis of this survey we will develop them.

Continue the systems that are in place

I really like our current solution. Our on-site NAS is very fast to access across the building network, so we store and analyze data straight from the NAS. The cloud based continuous backup is great, but accessing that data from offsite is slow.

Keep working with UCLA cloud data service (google drive since box is not available for large backups; some USB drives; duplicate NAS; if we ever get a insititution-provided network drive I'd use that

Instead of configuring backup, I will store important date on cloud storage.

None

Set up a database and create AWS or similarly-situated auto-DB backups

40 TB Server Raid 10 with NAS All student p-doc towers are RAID1 I provide external 8 TB Harddrives to students

Lab data on UCLA cluster (hoffman2) is already auto backed up for us, but we run out of that space quickly. The Synology NAS could be backed up to cloud but many options are expensive, so haven't set that up yet, still looking into it.

We plan to utilize cloud storage whenever possible, with annual back-ups from cloud storage to network drives

Use cloud systems

Remember to do it regularly

Maybe shift to Box

I don't have a good plan in place now. Backups are irregular and I realize we are at-risk.

We are happy with DGIT. They maintain our network. Also known as "IT Connect," or "CARE CONNECT"

No change planned.

Dealing with my backlog of things to back up to the server and external hard drive.

What future plans do you have for backing up the data in systems under your...

Google Drive via g.ucla.edu

Movement of data to enterprise systems that are hopefully backed up

Continue backing up to departmental server and local encrypted external drives.

continue to back it up in personal Cloud account. Talking to DGIT to create a server system at UCLA.

We have 25 years of experience in large-scale storage systems. Cloud storage is now by far the most cost effective method of storage. Over the past year we have moved all of our systems over to Wasabi and we have been happy with the service.

Our plan is to migrate most data to Google Drive and rely on their security.

Buy a bunch of external hard drives.

Would love it

I have none. But need to.

Same as current - multiple images of large datasets on different computers/externalHD

My research group may soon undertake more large calculations which will produce large amounts of data. Something like cloud storage may be needed.

I'm retiring.

I would like to move everything to a UC operated cloud server

I plan to continue to back up my personal computers to each other and to an external drive. This has proven to be more secure, robust and accessible than any cloud or server / cluster backup "enterprise solutions", which are much more likely to be hacked.

Large centralized storage solution in the range of 5 PB or more. No cost of in/out data transfers.

We are looking at moving to Amazon Glacier or similar for backup.

I would like to have cloud-based backup of our within-lab NAS system, but we don't currently have this set up.

Always have an off-site backup, try to outsource the the problem/responsibility to a cloud/service manager

I would like to see something that allows users to self serve restore, supports the major OS's like Linux, Windows, and OSX, and take into account that some devices like laptops don't sit in a single place for long periods of time.

We will continue to expand our local Eucalyptus footprint. It has proved more reliable that AWS or other public cloud solutions and far more cost effective.

Continue as I currently am, unless a more favorable solution becomes available.

None

What future plans do you have for backing up the data in systems under your...
buying another external hard drive
same as now
Purchase high density 50G Blue Ray discs
Cloud back up and redundant storage drives.
none, solution works for now
Everything is dependent on funding and what is available to me.
Get more of it into our current solution
Continuation of current practices
Preference for inexpensive cloud storage.
pending additional funding it will be difficult to greatly expand our capabilities
Adding an additional Synology NAS to sync them to high reliability availability. We use a Synology NAS for our server and two can be link to provide redundancy.
Just winging it like we always have.
anything inexpensive and easy to implement
None
Implement our tape backup solution
I have been buying computers and hard drives for my unit out of pocket.
Expand NAS capacity
An annual backup to an external disk.
none
External drives
Veeam
If I get a grant involving a lot of data, I budgeted for automatic backup to get set up.
I plan to continue doing what I'm currently doing.
Continue making archival copies of all data on offline, external storage.

What future plans do you have for backing up the data in systems under your...

Our current solution works for us. Data integrity may be affected by the frequent power outages on campus. Our UPS was destroyed by one of these outages.
More centralized system within the lab, rather than ones tied just to individual computers
Continue to back to removable drives
Continuing to backup data on cloud services + external hard drives
Moving more and more to TrueNAS Scale systems onsite and offsite.
rollout or enhancement of endpoint backup system, possibly also offering this as a recharge service
None
SDSC primarily backs up customer datasets on clusters, network drives, public and private cloud, and instruments. It's uncommon that we are asked to back up lab-controlled desktops, laptops, or other lab-controlled storage devices.
More AWS and Azure, more offsite backup options.
Increase off-site storage
Increase in onsite and offsite storage
na
Evaluating moving from tape backup to Azure backup.
na
na
purchase more commodity NAS (like QNAP or Synology) to use as backup targets
For research data, none really. We have discontinued the use of faculty using web servers to create their own website and moved towards cloud based solutions like CampusPress.
Want to integrate a MAMS/DAMS with existing storage and backup. Storage of tapes at multiple sites. Upgrade from LTO-7 to a newer technology at some point
leveraging AWS storage gateway
na

Q4.12 - What additional thoughts or concerns do you have about using a centrally funded or partially subsidized backup solution?

What additional thoughts or concerns do you have about using a centrally fu...
none
My research was funded by external Federal grants. Yet Berkeley provides no repository for data files, which seems wrong to me. The direct costs are gone; I don't have funds to pay for storage of the data.
Just cost. No funds available...
we should do it
We need more than just storage of the data. We need a way to organize the data for the 2 year proprietary period and then after that period the data must be searchable based on qualities of the data (meta-data).
My team uses Google Drive for most research data. Automatically backing up locally stored datasets on desktop and laptop computers presents the biggest challenge, because our organized research unit does not currently support it. So we're backing up on usb drives on an intermittent basis.
no comments
I would like to learn more about what this entails.
I would welcome a centrally funded solution if it was cost effective. My primary concern is that my backups are not far enough apart geographically.
I think core facilities (e.g. microscopy core, NMR, Mass Spec) should be the first and highest priority since we are institutional resources that are trying to provide professional level technologies to users, and yet we're operating on Mickey Mouse IT systems.
If cloud based, the network speed would be a factor.
None. Please create such a solution.
A properly implemented, robust solution with extensive version retention and a high intake rate would be a great asset.
Please let us be involved in selecting the solution (the opposite of FireEye implementation.)
Ability to control and configure backup needs to fit individual needs of our laboratory
The terms "centrally funded" and "partially subsidized" are separate possible solutions. Lumping them together is misleading for the purpose of this survey.
The cloud back-up solutions currently being promoted through IS3 are cost-prohibitive for individual research labs storing 100 TB of data. Wynton sounds like it will be cost-effective, but it is still being developed, so doesn't meet all our needs as yet.
My only main concern is if administrators are making the decision regarding what backup solution to get solely based on the responses of a survey. Such sorts of decisions tend to result in solutions that are economically or nepotistically convenient, but not at all practical for the purpose they're supposed to be serving.

What additional thoughts or concerns do you have about using a centrally fu...

1. A massive investment in a system that fails to meet the needs of users, but nonetheless takes funding away from academic units. 2. Security failures for systems chosen for the wrong reasons.

Past experience with centralization of services is that it generally is not as good as what we move from. This has certainly been the case with IT support in general. Plus I just dropped 20k for a bunch of this - the contemplated centralized "solution" seems like it is WAY too late and meanwhile I have to keep purchasing new machines for my research group.

Seems like a great idea to me. Obviously different folks will have different needs and for us, the data security is paramount. It can't be accessible from the internet.

Longevity of the plan. I don't want to have to move things again in two years.

I would use it if the cost was reasonable. $100 per TB per year.

I am disappointed with the backup solutions currently offered by my department. UCLA needs a centralized backup solution.

A flexible user interface is required, as well as modular architecture that would be robust to attacks in some parts of the storage structure affecting other parts of the system.

UC has had major data leaks over the years. I would be very concerned about a centralized solution.

This was part of what grant overhead funded at my prior institution, and I am in favor of a centrally funded backup solution, at least up to a certain allotment per lab. It would be simplest if this were just a large cloud storage allowance, but access to a networked department server that is remotely backed up would be fine too. Any solution should not prohibit labs from devising their own internal solutions.

It would be REALLY nice to have this option, especially since the university is taking away the main ways I backed up data previously by instituting quotas on Box and Google Drive accounts.

I think if there were a managed solution with comparable costs to a NAS or AWS storage, many people would use it so that they do not have to administer their own backups.

Longevity of the service, overhead in leveraging the service, and concerns about where the funding to support the service are coming from. We are already resource limited.

Access and installation. We had a departmental server in graduate school which worked great, but it was also set up on all computers (and could be set up on personal computer as well with ITS help). It was also nice because then instrument computers could access the data directly and save directly. If theres not access, it won't work. Also - cost is obviously a concern. Its hard to find funds, especially for new researchers before grants start coming in.

I've pasted together my own privately funded solutions because past institutional systems have suffered from poor accessibility, high cost, high administrative load, and poor user support.

prefer personal control over backup system

I would love to use such a system if it is not too expensive..

While backup costs can be included in future proposals, I have currently no funding to back up existing data archive (1-2PB) to commercially available cloud storage solutions.

none

Not very many people would use it. Hackers are very smart. I'd not trust any data storage on a campus server or in the cloud.

What additional thoughts or concerns do you have about using a centrally fu...

Centrally-funded has great potential for economies of scale, but may not work well for units involved due to administrative overhead and University controls. But storage is so much cheaper at scale that it could make a lot of sense. Cloud providers for storage are prohibitively expensive per Terabyte, especially when considering egress charges and penalties for accessing large chunks of data -- funding needs to be set aside if such charges would prevent restore. Also, network bandwidth for remote backup and restore will likely become a limiting factor for a centralized system for research data, which is so much larger than typical campus enterprise systems are built to handle.

limitation on how to access the data at any time, solution costs and most importantly speed of recovery.

I would love to see this happen. We should not need to keep buying hardware to admin machines for a few years and then replace it all (just to keep backing up the same blocks on different hardware).

Must accommodate large file sizes and large data uploads.

It should be compatible with our infrastructure - OS and hardware and minimally intrusive.

I will not spend a single second of my time in one more corporative solution which will only make my life more difficult

We really need more storage space on Hummingbird. Our lab is producing large amounts of genomic data that is too large to store anywhere else.

All students should be offered these back up services also

none

I think it's a good idea to have external solutions for backup that is independent of the researcher for cases where the researcher fails to comply.

I've heard that UCSC was going to lose unlimited storage on GoogleDrive, which would cripple my research.

We have systems in place to back up large data sitting on clusters. The raw data are also kept offsite, and so reproducing results is possible even if the data are lost. However, I am sure students and postdocs on my team are bad at backing up their laptops. That is maybe more prone to losing data and since most of that data is produced through hours of their own work, it is very difficult to reproduce. So I would try to make that very easy for the students. Starting with many USB drives that they can plug in at office/home (with appropriate adapters) would help a lot. That is probably better than online solutions unless the online solutions are seamless and don't have privacy issues.

Ease of use ought to be high

We run GITLAB on a virtual machine, and cannot get the use of "project disk" space because of the way they have configured the project disk and the permissions that GITLAB requires.

it is a good idea for everyone that does not want to use the cloud

the lack of support and information about switching backup solutions

Google has very good security. Would our central solution have good enough security? Or would we be putting all the data in a place that can then be attacked by ransomware?

it is certainly much better if the university provide a default back up solution for research labs, other than each of us figuring out that ourselves.

As far as ransomwear attacks go, wouldn't this potential solution be a very tempting target? How would you keep everyones data siloed so that my data won't get attacked if someone else is careless in this shared resource?

Everyone - including graduate students need access (permisison based) similar to that available with Google Drive

What additional thoughts or concerns do you have about using a centrally fu...

Privacy.

I will be out of my control, expensive and unaffordable in case of decline in funding and not easily accessible. Also, concerns about availability of qualified personnel for support and help.

In my opinion, the larger issues are determining scope / finding all of the research data, and assigning and funding backup administration ownership.

To make Linux service available.

Need it to be HIPAA compliant, FERPA compliant, easy to work on campus and off campus. Need it to be accessible from PC/MAC. Need seamless integration with Health Sciences campuses since clinical investigators need to be able to store research data there as well. There has been no solution for a central data storage system and only what we CANNOT use. This is unacceptable in a research intensive institution.

any solution(s) must meet CFR stipulated backup for Good Laboratory Practices requirements

As long as it does what we need we will welcome it!

Cost and convenience are the main considerations.

This would be great. It would be important to have enough help in the beginning and stable support long-term to ensure it is used effectively.

regular/routine access to data

Provided the solution can interface well with clusters, desktops, laptops, etc., I think this would be a big help to research on campus.

N/A

Speed of transfer, ease of installation

Security breaches

Data needs are very specific to research going on, and the heavy users undoubtedly have their own constraints. Subsidizing research work that is part of grants should not be done by UC.

I really think this needs to be offered. I have received no assistance at all from UC to make sure my data is safe and secure.

This is an increasing problem for us. We would love a central solution. E.g. a campus-negotiated solution at a cloud provider.

It could work as long as it is as functional and simple to use as Dropbox. Ie completely integrated with a laptop or desktops file folder system. Ie function invisibly.

I would be concerned that any central system would increase the vulnerability to a massive attack. It seems more secure to have a more distributed system.

nowadays we need a solution that we can use from off campus, and one that does NOT rely on internet connectivity; when I travel (especially when traveling abroad) sometimes I cannot access the internet

Would welcome such solution!

What additional thoughts or concerns do you have about using a centrally fu...

This is something badly needed by research labs - even basic information about standardizing, securing, and backing up data. This is something that should be funded by overhead/indirect costs and required by UC in the similar to lab safety services/inspections.

This would be greatly welcomed but would need to be a "set it and forget it" solution and come with educational support.

Integrations with a variety of systems and our ability to manage it.

Security, a central solution makes a lager ransomware target than individual solutions

Needs to not be too cumbersome to access data and run various analysis software, e.g, R, stata.

It would be great if UCB's data offerings were more affordable.

None

It won't be useful for my research group if it doesn't work with Linux in a standard way: mounted filesystem, sshfs, rsync, rclone, etc.

I am researching the

Fee to access rate structures should be explained explicitly.

As long as we can access the backups easily without restrictions - sometimes we realize we need a backup and need it rather quickly

Hopefully it will not be mandatory... I don't have large amounts of data, so I don't see any reason to be forced into a system that is designed for different needs

Please get us Dropbox

I am in support of this. However, given Google Drive is free, and unless I am required to use an additional data backup method, I probably wouldn't use it unless is was free and it was easy/automated.

Please worry about first providing storage for research at all, rather than having everybody fend for themselves to find the funding. Only once storage actually has a business model on campus, and an understanding what campus provides (i.e. more than nothing, which is today's standard), should we take about backup. Putting differently, today UCSD provides backup of 100% of the storage it provides because it provides none. ;-)

"You agree to indemnify UC San Diego against any third party claim based on an allegation that UC San Diego has violated that third party's copyrights, trademarks, other intellectual property, privacy, publicity or other legal rights due to your actions or omissions." is my concern.

A centrally funded solution should have redundancy and modular design to avoid a hack to a centralized system.

reliable fast internet connectivity is crucial, and not all UC Reserves have this capability.

The devil is always in the details - the idea of centralized backup to ensure data/system security sounds reasonable, but I anticipate researcher push back on: 1. cost overruns (many recent projects have ended up far more expensive than initial estimates), 2. tech support (will somebody who understands our operations/data/needs be assigned to us, or will anyone who is free answer our requests for help? Will we even be able to speak with someone within a few hours?), 3. scope creep (will this lead to mandates that make it difficult to perform research, particularly for small projects or for projects with minimal data/no sensitive data), 4. administrative hurdles (how many forms and committees will be involved, and how much time will it thus take, to be able to store/retrieve/move data), 5. on-going access (if I move to another university and continue to collaborate or am finishing up projects, how difficult will it be for me to access the data).

Achieve economies of scale on licensing and support units that have data with regulatory compliance needs (HIPAA, CUI, etc.)

What additional thoughts or concerns do you have about using a centrally fu...

Major concern is that cost would be higher than owning the storage hardware

I am worried about individual laptop users in the arts and humanities. Losing a manuscript is career altering. There is a sense of complacency that running anti virus and backing up to a external drive is sufficient.

We are happy to be using UCSF RAE for data storage and back-up, but wish that back-up solutions were more transparent and lost files easier to restore. Most of our work uses EHR data but we also need to make sure that complex code for extracting, cleaning, coding and analyzing data is backed up and easily recoverable...in addition to the analytic datasets (which are reproducible as long as the code isn't lost).

There is constant staff change in my current institute, so I am worried about the disruption of service if I were to switch to unit-controlled networks and drives instead of cloud-based solutions.

Major concern is who owns and has access to storage of government, military, or commercially sensitive data

Think centrally funded backups are critical to researchers who do not have a unit solution.

We are relying heavily on BOX as a secure back-up for our research data. It would be a major problem for our research if data stored on BOX were to become unavailable.

centrally funded would be good. I would still back up as an extra safety by other means.

I need a solution within the next 6 months. If a solution can be established in that time frame, then I am very happy to buy in.

Need more information.

It's been difficult to use some of the systems that the university supports, like Box and SharePoint, because they don't allow for the customization that we need (Box) or they are very user-unfriendly (SharePoint). If there is going to be a centralized backup solution, it needs to allow for customization, be easier to use, and allow researchers to not be required to use it.

We need to be able to easily access backups, again, this is MS SQL Server files, and sometimes we need to look at older data.

This is a huge need in the research community. As the director of a core computational genomics facility, we work with >50 labs a year processing their big data. Most of these labs do not have the storage capacity or knowledge to back up their data to their own servers so we have a data storage service to help them, but this is not ideal, as each lab should have access to a secure, backed up storage solution that doesn't have a high cost and is also not the old dusty external hard drive in the bookshelf. In addition, I have the perspective of a new faculty member just starting up a lab and looking for data storage solutions. I explored the San Diego Super Computer Center, which sounds like a good option, but it is almost 2x as expensive as AWS S3, so I am currently planning to do my data backup on a new Health IT managed AWS account.

Our research data is also administrative / operational data, all collected and managed by the system-wide UC Recruit service. We are probably not your typical use case but are interested in this initiative.

I think it is a case of a few having problems and trying to get the rest to bear the burden. I DO think, however, that having UC guidance to researchers, and recharge fee options, are a good idea. But all researchers have, or should have (it si already mandated by NSF/NIH) data management budgeted into their grants.

For most users I suspect will not be practical if there is cost associated with it (grants run out and can't keep paying fees to store data after grant done), bandwidth would need to be truly 1 (ideally 10) Gbps (currently may be limited by campus network also and not just storage server.)

None really, it depends on what the specifications are for the solution provided.

This needs to be centrally funded completely since this is what indirect costs are for.

What additional thoughts or concerns do you have about using a centrally fu...

I would welcome such approach. I thought we had with Box, but that seems to be going away.

data privacy and intellectual properties.

Current solutions are far too expensive. We do not get any additional funding from sponsors for this and the list of requirements is always increasing. We need something that can slip in without too much funding to maintain an offsite backup. Google Drive used to work, but is no longer available to us.

We may be unwilling or unable to fund our portion of a partially subsidized backup solution.

i have none.

UC fails at nearly every digital thing it does, from being hacked repeatedly to being saddled with enormous expense for user-hostile solutions (UC Path). How can you, at this point of complete and persistent opacity regarding the massive recent hack that exposed personal information of employees, turn around and ask the same employees to envision a system that they could trust?

1. Speed and convenience of data access are paramount. 2. Not reinventing the wheel, but rather using robust, mature, well supported software that has stood the test of time (e.g. rsync). 3. Funding should include support for multiple backups (at least 2). Ideally this would include one backup copy on physical devices in the lab's possession and a second master copy on a remote network and or cloud-based system.

I am very worried about costs -- what incentive will researchers have to keep backups reasonable? I am also very worried about cumbersome interfaces or APIs that do not meet our needs or interface with our existing technology.

Ease of access compared to existing solutions such as Google Drive

I am concerned this will end up being just another survey with no financial commitment from the institution or follow up.

Still a significant amt. of work on desktop/laptop due to difficulty working smoothly in the cloud (box/drive). Upload of data/final product is a chore and therefore not always done. UC computers should be automatically backed up. In my case, I chose an IT-managed computer and was told they make an image and then my hard drive broke and it turned out they hadn't backed up anything and is was a miscommunication. We should have centralized default backup for all UC-associated computers. If users have to set this up for themselves, the failure rate will be immense.

The university should simply provide materials, migration, and access to storage but allow researchers to retain absolute control over access of the data

I have had over the years a continuous process of upgrading backup systems based on evolving hardware and software, and especially availability of personnel who can support the effort. For example, after over 30 years, I have gone through multiple generations of storage devices and media, and the simple process of migrating to the newest generation for numerous computers is getting to be a great burden.

overpriced and over complicated solution offered for simple backup needs of research, computational lab

The security of the centrally-funded solution would be paramount

I think we need to move to cloud, if DOD and Pentagon are, how can we not?

This is a great idea. I already move my work files to OneDrive provided by the campus and it works great.

Who controls it, and what empty promises come with it. If it's anything like the ESR rollout, we will not want any part of it.

We definitely need more and better solutions for routine back up of research data.

What additional thoughts or concerns do you have about using a centrally fu...

Good luck.

Finance model is key -- costs have to be "just right." If you set the cost too high, people will be driven to other more cost effective solutions and no one will use it. If you set the costs too low, you risk becoming a tragedy of the commons data dumping ground where no one has incentive to manage or control what their storing and costs of providing the service spiral out of control.

None. Sounds like having a uniform solution and process and system is better than every person acting piecemeal with different media.

Centrally funded will work. Partially subsidized will fail in the case of research because research projects are often run on a shoestring budget.

I would continue to back up my data in the ways just described even if there were a centrally funded or partially subsidized backup solution.

As long as we're able to design a solution to fit our needs, we'd welcome financial support. :)

If only partially subsidized it will be more expensive than our current low cost solution.

Would love to have a clearer idea of storage quotas moving forward

I love this idea! Wish I didn't have to figure this out myself.

I was surprised that you didn't ask any questions related to what OS or filesystems the data is stored on. Some research groups use Linux to various degrees, but presumably you folks know that.

Really our only concern is cost. Central solutions that have been offered to date tend to be priced at $/TB/month at a point that is vastly more than the $5K or so per year on average we spend to replace or upgrade commodity SAS/SATA drives or enclosures to support our (currently) ~200TB backup needs. I believe we were quoted $40K/year for Rubrik; in comparison in 2021 to date we have spent $200 on our existing solution. If a "free" central solution were provided, we would certainly use it.

Great idea, as long as it remains flexible and alternative options can still be used/integrated/reimbursed/paid for by lab funds

We are price-sensitive Servers use zfs filesystem and use zfs features (snapshots, incremental streams) for incremental backup. We would want a solution that can interact reasonably with zfs and the client is not too intrusive.

This may generate unnecessary issues with the research.

This survey seems to be directed an people who know what an entire unit is doing. I do not. I am merely an individual in the unit. I think, unless I misunderstand, that this survey should only be directed to people with knowledge of their entire unit.

I have no problems with a centrally funded or partially subsidized solution, as long as it is affordable for research labs and easy to access/share/control as Google Drive has been.

We deal with large audio / video assets that consume lots of storage space. We have been running our own QNAP NAS because it is cheaper per Gigabyte than the campus services we have investigated. Having recently suffered a ransomware attack, we recognize that we are not equipped with sufficient staffing to manage this size of archive without loss. However, the high price per GB of campus managed storage has forced us into this position.

Price and security

I have zero confidence in UC's capacity to provide a solution I would trust. UC is a target; my lab is not. UC has many priorities; my research/data is one of my top priorities. I am committed to protecting my data at all costs. I do not expect UC to share this priority.

What additional thoughts or concerns do you have about using a centrally fu...

There should be two solutions: 1) Centralized storage for rapid recovery of loss in case of local machine failure, 2) Offsite backup (Backblaze B2 or similar) with a data retention scheme that allows recovery in case of campus-wide failure. I would also suggest that for (1) above, we have agreements with other campuses, to allow retrieval of backups from the U.C. system. This would reduce the cost of recovery from the cloud in case of a campus-wide failure.

For data on our own computers (which is not much), we use Crashplan. For data on UCSF systems, I assume/hope it's already regularly backed up! I think it's great to have it centrally funded.

Compatibility, ease of administration and use.

Cost, access.

The charge. I feel pretty strongly that all of my lab's computers should be able to be backed up easily and without charge and I feel pretty strongly that all of the data that are not stored on our computers should be able to be backed up easily and without charge.

Interruptions to my research activities.

Privacy and continuous access

A lot of North Campus faculty have data that's valuable/irreplaceable on the cloud or laptops, and because most of the campus focus is on bigger data, labs, and the like, this probably falls through the cracks. I'm not sure of the best solution.

There is an enormous need for backup solutions that can handle sensitive data appropriately. Ideally, encryption could be client-side using a researcher-controlled key.

We would be concerned about costs, service interruptions, and network speed. Speed is critical because we are frequently moving TBs of data.

There are several systems currently available to UCLA researchers whose cost prevents them from being used on a typical research grant budget.

A system-wide backup solution must be flexible and under some form of local (research lab) control and not administered by campus IT which has a poor track record of understanding the needs of researchers.

I have worked with IT to setup cloud storage. It took one year to get it set up, was clearly not ready to roll out, and it is still difficult to access and transfer data (frequently kicked off cloud storage server). We desperately need to catch up on a solution.

For us to use any solution, it would need to be compatible with multiple operating systems (RedHat Linux and Windows Server for us). Since much of our data is static (at least 20TB is from 30 years of observation data), a one-time local copy solution (USB or other) for the bulk would be important, rather than trying to transfer all to backup via the network (which would potentially interfere with our realtime data collection).

Other than centrally funded solutions, one time costs are the only realistic option for grant-based research.

I don't want to have to update it periodically, I'd like the same solution to be available over several years.

Needs expert oversight

Easy of use and throughput

Low cost is critical, but also the functionality and ease of use both for the backups and the restores.

What additional thoughts or concerns do you have about using a centrally fu...

None. We'd be so very very happy to have this.

Our group handles human-subjects clinical research data that is subject to numerous regulatory requirements regarding data security and privacy. Many projects also involve data collection at Department of Veterans Affairs sites, therefore we will only be able to use systemwide data backup solutions to the extent that they conform with VA requirements (e.g., VA-approved encryption and access control options).

It has to work and be easy to use.

Provide research groups with AWS credits and technical support to implement a robust backup solution (hot vs. cold storage) that suits their specific needs. Do this - and everyone will be very appreciative.

Amount of cost, privacy, support

Seems like a good idea although you'll never make everybody happy! I've learned the hard way from administration that you'll need to communicate over and over again and also do a huge amount of training and hold people's hands through the process. It is critical to make sure the benefits are clear (to the PI, not just the university, as many people are selfish) and to minimize the transition costs.

The cost is the main concern. We would prefer a solution that does not increase our costs.

That putting it all in one place makes it more vulnerable

UC-wide or campus-wide solutions almost always become less than ideal. I think the best is for the UC or campus just "help" individual PI's find a commercial service. If UC can negotiate discount prices, that will be great.

Ease of use, privacy concerns,

My concern is that the support will be poor or non-existent and too expensive for the service offered. The email service out of sub-units on campus (e.g., SEASNet) is bad. Webhosting service for our lab websites has been offered but doesn't seem to be professional enough to justify the cost even partially subsidized. UCLA Box is too slow, has a poor user interface, and overall is disappointing compared with Google Drive. UC Path system has had a disastrous launch with bugs that lasted for months. Overall, UC and UCLA's track record in IT Service is not great! I am concerned that a partially subsidized back up solution will be similar: too expensive for a poor service with inexistent support.

none

UCOP mismanaged the Accellion FTA incident and is still not disclosing what happened; see, for example, Sindhu Ananthavel's report last month <https://dailynexus.com/2021-09-05/uc-refuses-to-disclose-information-regarding-accellion-data-hack/>;. I worry that any UC-wide backup system would have similar issues.

this needs to be funded -- I have used personal funds for my solutions, yet my job requires that I have adequate, convenient, fool proof solutions

hacking, virus, security issues

Anything partially subsidized will cost me more than my current $0 UCLA Box, cloud-based solution.

None

More bureaucracy, more $, less service

Centralized backup solutions need to be HIPAA compliant.

What additional thoughts or concerns do you have about using a centrally fu...

One of the problems that makes researchers reluctant to rely on Box is the frequent loss of data, mismatch of file types, and other general errors that occur with box. Many of the researchers feel it is easier to manage their own backups.

Please give us a cloud storage solution.

Unfunded mandate that would further cut into limited soft money salary.

Certainly need to have automatic backup daily

trust. what is promised up front in terms of security, accessibility, reliability and cost is not necessarily what is sustainably delivered. If the terms change, or promises aren't kept, and we have not invested in our own solutions, we cannot deliver on our commitments to sponsors. They are not interested in the reason why. The personal reputational/career damage can be very high. Given our experience with University-run services, it's not a risk we would lightly take. We would probably only use a university service if (1) it were free and (2) it was for longterm offsite duplication where access was not a mission critical dependency on short time scales. The University of California cannot be trusted where IT reliability, accountability, accessibility, sustainability, design, etc. are concerned.

It should be at not cost to the PI/lab, or at least for a minimum "decent" storage (few terabyte per lab personnel). Each researcher should be responsible to back up his/her data with a direct and all time unlimited access.

If cost is reasonable, I would embrace it.

I am not the right person to respond to this survey. The data I have is so small that elaborate systems are not needed.

No concerns.

The cost and the IT support to use it appropriately.

This seems like a great use for the Technology Infrastructure Fee that is added on to every grant.....maybe increase that to pay for this.

Lack of control over timing of backup; need for IT assistance in retrieving backup data

It should be simple, easily taught and disseminated, and not geared totally to the hard sciences or medical schools.

Research data management, storage, back up should be supported by the institutions, as part of research F&A costs. We are already paying for TIF as direct costs.

Needs to be easily accessible from different work places (home, office, during travel)

My only concern is access speed/bandwidth. We currently have ~20TB capacity on our NAS, which is what we need.

That it actually happens. We've had a need for this for 20 years.

It has to be 100% secure

None

there is NO IT support. I would love mountable network drives, but they need to be reasonable and there needs to be some assistance when things go wrong

What additional thoughts or concerns do you have about using a centrally fu...

Flexibility to remain useful while my researchers are in the field, often in locations without strong internet connections. Integration with long-term preservation services that facilitate reuse and reproducibility of research. Sufficient personnel support to ease adoption and maintenance over time.

I was really surprised that no such solution exists for UCLA researchers

I would welcome a centralized solution.

I would welcome an automated data backup option for our laboratory research data and regulatory documents. However, I have concerns that recent centralization of certain systems in the UC, and local centralization of IT in our department, has created substantial backlogs of service requests and generally made things harder to accomplish in a timely manner. Moreover, F&A on extramural funding already includes a lot of line items and I would be hesitant to take advantage of a central back up option if its cost were prohibitive. To me, the ideal backup solution would be automatic on a regular basis and easily accessible.

We do not have money available to start using a new backup system. UC should provide backup. Like they are currently doing with DGIT.

I don't understand it well enough to have an opinion.

It would have to be free or my PI would never agree to pay for it.

Google Drive via g.ucla.edu should be emphasized for the majority of campus faculty. Only large research data should use other solutions.

One significant reason why centralized solutions fail to meet researcher needs at UCLA is inadequate communication about how a general solution applies in a specific context.

We can't currently rely on the university system (departmental server or BOX) as sometimes the connections go down and server storage is not completely stable although generally very good and has improved. Box always confuses between my emails and logins, and sometimes university connections time out if traveling in internet-poor areas. These are improving but stopped trying to use them when there didn't meet needs.

so many different needs

Right now we are using Box almost 100% - I am concerned about data loss if something happens to that cloud service but that is the best solution right now. We should ideally add a backup service to Box that is separate.

I would be happy to use a UC solution if it were as cost effective as the commercial providers. However, in that case the campus should also pay attention to network speed, because that is important when storage is remote.

It needs to be seamless and automated; anything that requires frequent human intervention is not likely to be maintained.

Increasing costs, opaqueness of management, possible failure of the solution (if it doesn't have backup).

Would love it, don’t have time to handle alone

That it won't be user friendly, and will require significant rollout effort by the UC to implement.

We would love to have a system that was easy, secure, but most of all, extremely inexpensive.

I do not need it or foresee using it. My main concern is that "enterprise solutions" are much more likely to be hacked than personal computers. It seems like an expensive solution to technically challenged and / or lazy people.

It needs to be big enough, generating 3 - 5 TB a day with sets of data exceeding 10TB can make you anxious about where to put everything. Box was slow but useable. 112TB on campus provided storage is a bit too small.

What additional thoughts or concerns do you have about using a centrally fu...

Ease of use, ease of data recovery if the worst happens and our system goes down.

IT support to work out installation and network security problems related to the backup will be crucial. Otherwise, we will get instructions on how to use the solution, but some fraction of computers will have an install problem and we cannot assume that lab personnel know how to overcome these themselves.

At UCSB we also desperately need a way to store PHI/PII/HIPAA data!

Everything costs money, and if there is a campus, or unit provided solution that takes into account the cost of doing business rather than shuffling UC money between departments it might be worth investigating. I usually wait and see what is being offered before making any decisions, because some sound great, but end up being junk on the software side and require you to be a programmer to use. I think any system should be subject to a hands on demo from the people that actually have to work with it as the pass fail before buying in.

Cost and ease of access are the primary considerations. Currently, our costs are close to zero as Eucalyptus is close to self-maintaining and the access uses the AWS APIs (which is familiar to the research staff). If we were to move to a centralized solution it would need to be similar in cost to our group and ease of access.

Backup seems like a wonderful aspiration but feels herculean from my vantage point given that data are all over the place on machines and servers that are not supported by the university. Specifically, there are currently lots of incentives to implement faculty/individual/personal based solutions for data storage and backup that exist apart from the university's services. One example is the understanding that our school's computing policy allows support/services for university-supported purchases of equipment (but the sticking point is that these equipment cannot have us researchers be administrators on our own machines; therefore, I don't know of any faculty researchers in my department who regularly use laptops/desktops supported by the university). The idea that we'd have a backup solution seems at least somewhat dependent on the assumption that faculty are already supported by the university for data storage/computers/etc., but that isn't the case in my school.

that centralizing the data would mean that if someone could hack into that system, they'd have data from everyone and cripple the whole university in terms of research

In my mind, the limiting factor would be the cost that the faculty member would need to pay.

need staffing to help set it up

it is costly to keep adding NAS space and we keep hearing that the cloud storage will soon cost us too. grants are 3y long on average but much of our data is supposed to be kept for up to 25 years per funding agency guidance

Some guarantee that it would exist for a decade or more. The backup we did to google took us about 2 years to complete only to leave us in a lurch. Being grant funded almost excludes any sort of recurring charges.

Longevity of the solution and scale. Need many PBs of storage

Creeping costs and inconvenient migrations in the coming years.

Thought should be given for combining this with repository to make data public.

Please make it happen

I am concerned that Google may limit our backup in the future. This is a a UC issue as UC negotiates with Google as to limits.

Main concern is it just disappearing someday because it gets too expensive (like google drive) or dumping key features like linux support (like dropbox).

Off site storage would be an additional layer of protection that I would benefit from. As I have confronted my own storage issues I have also faced the reality of the slow network speeds in my building.

What additional thoughts or concerns do you have about using a centrally fu...

There needs to be some assurance that it will stay in place. A system that goes away after a few years and leaves me scrambling to put something else in its place is worse than no central system at all. We already had a scare with Box so the idea that anything centrally-provided can be centrally revoked is always on my mind.

What I'd really like is to be able to shift one of my backups to off-site. As-is, I can spread across rooms, but if the building burns down, I'm in trouble.

encryption at rest

We work across units. It would be very difficult if different schools or units or campuses used different solutions. I would like to use the google (or microsoft) suite. My field requires data backup on servers, so anything that makes it to publication is achieved at that time.

Historically, UCI support staff have been very rigid and not particularly responsive, going for lowest common denominator and typically answering many reasonable questions thoughtlessly and with "that can't be done" when it can easily be done.

Central systems might take too much time to bother with.

none

None

Reliable technical support is needed at no cost or it won't happen.

The problem has been many years in the making, and it will take years to implement a central solution.

We have some HIPAA-protected medical imaging data. These must remain under lab control. I doubt that backup for our other data is feasible, given current network speeds on campus.

Great idea - we would need to roll out enough information about it to individual departments on each campus so that faculty 1) know the option exists and 2) know how to use it.

Individuals should have full coinrol

It would be good if a central solution offered a public facing data repository with direct access URIs and a preservation funds so data can be made accessible in perpetuity

Reliability is all that matters here for us.

experience with central unit making arbitrary decisions based on the needs of their internal customers which may differ from researcher needs

needs to have reasonable download and upload speeds for large image files that will be collected on equipment in cores and analyzed on personal/lab computers in a different location Cost is an important issue

How well does this work cross platform, eg Linux, Mac, Windows, how accessible is the data, would there be provisions for cheap long term slow storage vs "faster" but more expensive easily retrieved storage (for items in more use / less use), how safe is the data from others on campus using the same central backup, how clear are the protocols, who is administering it and how available would they be for troubleshooting AND being familiar with multiple platforms, etc.

Recovery time, data transfer speeds, who has access to backup data, ease of setup, clients for each operating system (including Linux), would external administrators have access to view content?,

What additional thoughts or concerns do you have about using a centrally fu...

Environments that generate large volumes of data (many small files or very large files) through instruments/sensors and computer simulations present special challenges for storage/share/restore operations and may not get the attention they need with a single centralized solution.

We just transferred our 100Tb of data between AWS accounts. This was not an easy or straightforward task. Transferring data back to a central backup solution would need to be easy and provide validation that all files were copied accurately. Also, we would not want to maintain the hardware. Also the backup storage should be expandable.

I would welcome it if it was optional and offered with no cost to division.

na

It would be great to have a solution that works directly from researcher laptops/desktops -- much as we did with the former Crashplan option.

na

na

I imagine that law school researchers' needs shouldn't really drive a push for a centrally funded backup solution, but depending on the details, I imagine there are many researchers who might take advantage of such a move.

Encrypt our backups!

I can envision a system of ANY size becoming oversubscribed, and likewise I can imagine unwelcome restrictions on what can be backed up as a result. Also concerned about ability to browse OUR content and restore quickly as needed. And I would want assurances and mechanisms to ensure OUR data could not be accessed by other units.

access for retrieving data, transparency around the platforms scope and usage

na