COOPERATIVE APPROACHES TO RESEARCH DATA SERVICES

Melissa Cragin
San Diego Supercomputer Center
UCSD

Framing

- Active data – focus on services for researchers engaged in the research cycle
- Action leaders seeking both structural and process solutions to improve access and use of data within constrained budgets
- Faculty are interested – or at least willing – to learn about the trade offs and participate in planning around their needs, supported by
  - *people who's role it is to help train, facilitate the planning process, deliver services*
  - *transparency around costs and trade-offs*
- Sustainability requires some commitment from unit leaders or campus admin

Linked facilitators	Research Unit	All-campus Coordination	Institutional Investment
Continuum of service	Sustainable, flexible shared cost models	Reduce stress across units, improves trust	Data lifecycle as driver
Campus IT and SDSC provide info, “hand-off,” and consulting	Storage & compute - 100% access only during award - Dept. pool / shared - Campus pool / 50%	- Formalized Storage Council - Reps from all units that provide storage - Awareness of policy changes	All faculty, staff, students - 100GB private - 3 yrs - Publish 1GB – 10 yrs Funded projects - 1 TB private – 10 yrs - Publish 10GB – 10 yrs
- Range of “free” options on campus - Consulting - Infrastructure at scale	- Consultation - Genbank submissions	- Storage Campus Champions to id. best solutions	- DMP support - Publishing support
Unfunded	**Fee-for-service**	**Unit staff time**	**Administration funds**

Data Storage Finder

Evaluate options for data storage at Cornell

All services presented on this finder tool are vetted and supported by Cornell University.

To explore data options available to Weill Cornell Medicine Cornellians please visit the WCMC storage wizard.

We welcome feedback on this tool.

Describe your data

Answer these questions to help identify data storage services that are suitable for your needs. Checking these boxes will change the list of available services.
If you are uncertain how to answer, leave the question blank to maximize your resulting options.

Clear Answers

1. What is the classification of your data?

- [ ] Public
- [ ] Sensitive / Moderate Risk
- [ ] Confidential or Restricted / High Risk
- [ ] HIPAA-Regulated

2. Do you need backups, snapshots or replication of your data?

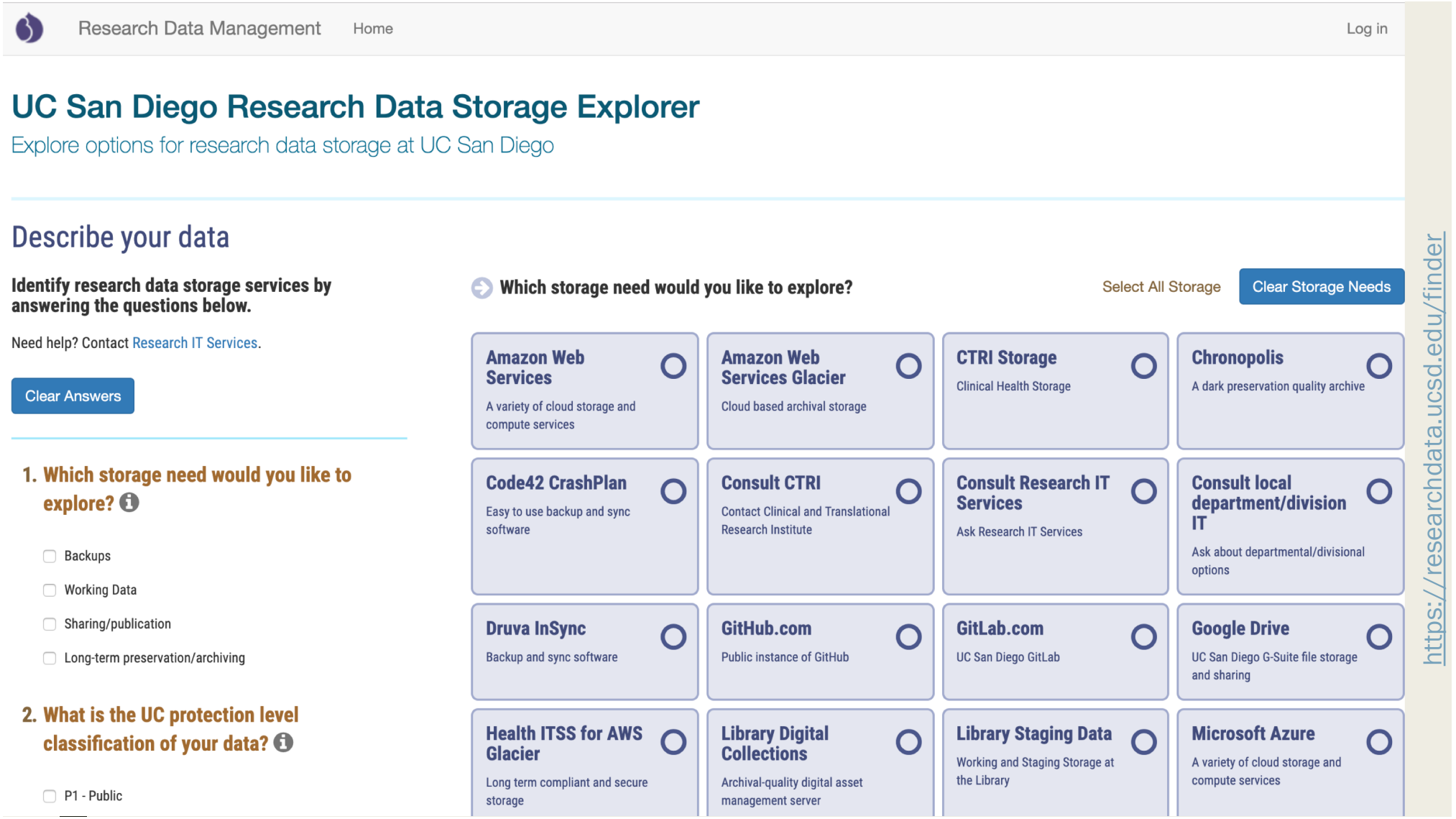
- [ ] I need one or more backup/snapshot copies of the data,

Select data storage services you would like to compare.

Select All | Clear Selections

Amazon Web Services Elastic Block Store Storage for use with Amazon EC2	Amazon Web Services Elastic File System Storage for use with multiple Amazon EC2 instances	Amazon Web Services Glacier Cloud based archival storage	Amazon Web Services S3 Flexible, scalable object storage
BioHPC Cloud Storage for BioHPC lab computing services	CAC Archival Storage Single copy, non-mountable storage	CAC Red Cloud Storage Storage for Red Cloud compute instances	CISER Research Servers Storage for CISER computing environment
CUGIR Publicly shared geospatial data storage	CUL eCommons Publicly shared data repository	Cornell Box Online file sharing and collaboration	Cornell Restricted Access Data Center Storage for CRADC (confidential) computing environment

https://finder.research.cornell.edu/storage


Research Data Management
Home
Log in
UC San Diego Research Data Storage Explorer
Explore options for research data storage at UC San Diego
Describe your data
Identify research data storage services by answering the questions below.
Need help? Contact Research IT Services.
Clear Answers
1. Which storage need would you like to explore?
Backups
Working Data
Sharing/publication
Long-term preservation/archiving
2. What is the UC protection level classification of your data?
P1 - Public
Which storage need would you like to explore?
Select All Storage
Clear Storage Needs
Amazon Web Services
A variety of cloud storage and compute services
Amazon Web Services Glacier
Cloud based archival storage
CTRI Storage
Clinical Health Storage
Chronopolis
A dark preservation quality archive
Code42 CrashPlan
Easy to use backup and sync software
Consult CTRI
Contact Clinical and Translational Research Institute
Consult Research IT Services
Ask Research IT Services
Consult local department/division IT
Ask about departmental/divisional options
Druva InSync
Backup and sync software
GitHub.com
Public instance of GitHub
GitLab.com
UC San Diego GitLab
Google Drive
UC San Diego G-Suite file storage and sharing
Health ITSS for AWS Glacier
Long term compliant and secure storage
Library Digital Collections
Archival-quality digital asset management server
Library Staging Data
Working and Staging Storage at the Library
Microsoft Azure
A variety of cloud storage and compute services
https://researchdata.ucsd.edu/finder

Some additional thoughts

- Affordances:
  - *Campus-based cooperative arrangements are on the upswing*
  - *Access to expertise*
    - Trend toward professionalization of research staff roles
  - *Increased numbers of people and projects are being supported while managing costs*
  - *Data as university asset – faculty recruitment*

- Challenges:
  - *Variation (institutional) of kinds and extent of services (e.g. no overhead for grant-funded commercial cloud cycles)*
  - *Hidden costs for data services across the lifecycle and service groups*
  - *Differences in procurement processes for multi-institution infrastructure projects increases costs and necessitates much higher management overhead*
  - *Lack of published empirical data on emergent trends and models*
    - decision processes to develop and run these

Acknowledgements

- Brian Balderston - Research Data Services, SDSC (UCSD)
- Lisa Johnston - Data Curation Network (UMN Libraries)
- Wendy Kozlowski - Research Data Management Services Group (Cornell Libraries)
- Brett Tyler - Center for Genome Research and Biocomputing (Oregon State University)
- Jim Wilgenbusch - Minnesota Supercomputing Institute (UMN)
- Michael Witt - Distributed Data Curation Center (D2C2) (Purdue Libraries)