Answering your questions

JAMIE WITTENBERG
RESEARCH DATA MANAGEMENT
WITTENBERG@BERKELEY.EDU

Free Tools?

	Best suited for	Key capabilities	Not well suited for	Limitations	Data protection/ security	Cost	Default allotment	Connection method
UC Berkeley-provided (directly or via contract)								
Box	• Parking/gathering data prior to or after compute • Collaboration • Stewardship of files • Off-site protection copy	• Unlimited storage • Share with colleagues outside UCB • Robust file management, roles & permissions	• Back-up High-speed data transfer	• 15 GB maximum per file • Slow file transfer	PL1	Free	Unlimited	Web browser, mobile app, sync, desktop tool, command line, API
Google Drive	• Parking/gathering data prior to or after compute • Collaboration • Stewardship of files • Off-site protection copy	• Unlimited storage • Share outside UCB • Robust roles & permissions	• Back-up • Serving as an end-point for high-speed data transfer	• 5 TB maximum file size (i.e. per file) • Slow file transfer	PL1	Free	Unlimited	Web browser, mobile app, sync, desktop tool, command line, API

More free tools?

Jupyter Notebooks (awesome. Document and execute code & viz in one place): http://jupyter.org/

Coming soon: campus Github! (Private accounts)

Free Programs/ Services?

Research data management (everything): researchdata.berkeley.edu

D-Lab (mostly social science, qualitative methods): http://dlab.berkeley.edu/

BIDS (lots of grad students, postdocs): https://bids.berkeley.edu/

Hacker Within (coding, computational methods, informal)

Data Acquisition and Access Program (Library): http://guides.lib.berkeley.edu/data

Berkeley Research Computing (HPC, big data transfer): http://research-it.berkeley.edu/programs/berkeley-research-computing

What amount of data merits ‘management’?

32 bits

2 petabytes (1,000 terabytes), continuous

It’s not just ‘amount,’ it’s type

Does UCB have tools to facilitate government-mandated data management systems / protocols for research groups?

DMP Consulting: researchdata@berkeley.edu

Easy to use database framework to store data?

Contact Berkeley Research Computing: http://research-it.berkeley.edu/programs/berkeley-research-computing

Best practices for making/keeping backups?

3-2-1 Rule (Anna talked about this)

Sync for Box & Drive (but not on collab folders!)

Backup vs. Preservation – We have an institutional data repository (and you get a DOI for publications – dash.berkeley.edu)

Metadata & data description best practices?

DDI (Data documentation initiative) - http://www.ddialliance.org/

Codebooks (aka what does that variable mean)

Data dictionaries (relationships)

Self-documenting code (http://c2.com/cgi/wiki?SelfDocumentingCode)