BEST PRACTICES FOR MANAGING YOUR DATA

To assist you in determining the best approach to managing your project's data, Engineering IT, in coordination with the Research Data Service, is providing this document which is intended to help establish a common vocabulary and encourage discussions and considerations around what might best fit your needs and to help you tailor solutions to your research project as a whole.

About the Research Data Service: The Research Data Service team offers expertise and guidance on many areas of research data management, from data management plan reviews, workshops, and consultations. All Research Data Service consultations are free and confidential to all members in the Urbana-Champaign campus community. researchdata@library.illinois.edu / http://researchdataservice.illinois.edu/

NOT ALL CLOUD SERVICES ARE THE SAME

While Box.com, Dropbox, and Google Docs are some of the more commonly used cloud storage providers used by the campus community, the University of Illinois has several cloud resources where they've negotiated better services to be available to faculty, staff, and students. Take the time to become familiar with their usage terms.

BOX - Illinois has negotiated 50 GB of storage on Box.com for all Students, Faculty, and Staff, but that can be increased if needed. Box folders can be shared to anyone with an email address - on campus or beyond. Box has a 15gb individual file size limit. Box is the only cloud storage option currently cleared for FERPA data and it has received approval to store IRB data. Box has an agreement with Illinois to prohibit the harvesting of data on Box folders and files, and all University of Illinois data stored on Box is stored on servers located in the USA.

GOOGLE DOCS / GOOGLE DRIVE - Google is another cloud storage provider for all Students, Faculty, and Staff and it can be useful for real-time collaboration allowing several team members to simultaneously edit a document via a web browser. Google Docs only works this way with a limited number of file types (Word, Excel, PowerPoint, JPG, and PDF). Google offers unlimited storage, with a maximum single file size of 5 TB. Google is not FERPA compliant, Google cannot guarantee documents stored on their systems will remain on servers in the USA, and Google has not signed an agreement with the University of Illinois to prohibit the harvesting of data on their service.

MICROSOFT ONE DRIVE - Microsoft One Drive offers 1 TB of storage to all Students, Faculty, and Staff. The individual file size limit is 10gb, and integrates tightly with Microsoft Office. Microsoft One Drive is not FERPA compliant, but in most other respects it has the same features as Box.

DROPBOX - Dropbox is neither provided nor supported by the University, but the way it operates is very similar to Box. Dropbox key differences include an initial 2 GB of storage and the fact that they have not signed a non-disclosure agreement with Illinois to prohibit the harvesting of data on their service. Dropbox is not FERPA compliant, and Dropbox's terms of service include a clause by which they can suspend or terminate services at their discretion, and without notice. Dropbox Pro is available at a cost of $99 per year for 1 TB of storage.

A more complete list of storage options and features can be found here:
http://it.engineering.illinois.edu/services/file-and-data-storage

IMPORTANT CONSIDERATIONS FOR MANAGING YOUR DATA

- **The way you preserve and index your data will affect your ability to use and recover it later.** For example, information in an individual user's Box folder may be inaccessible if they leave campus.

- **Strike a balance between consistency and uniqueness** in your file/folder names. Include important metadata (date, location, version number, etc) in the filename when possible.
  Example: 2016-10-19_MRL_Hitachi_S4800_Scans

- **Use the YYYY-MM-DD format for your dates** in all places, and enforce this in your labs, project groups, etc. This can help keep files displayed in chronological order, even over several years.

- **Avoid spaces and special characters in filenames**, as they can be problematic with various systems and software applications.

- **Discuss and document standardized or controlled abbreviations, acronyms, etc.**

- **Create project documentation**, even if you only have time to make it brief. Develop templates for your lab to fill out for every project and follow domain-specific metadata schemas when appropriate. One way to do this is to **make a readme file in each folder** to briefly describe the folder's contents, naming scheme, and acronyms.

- **Keep your raw data protected and unchanged - never work on the original.** At a minimum, keep an extra copy of the original raw data somewhere in case of accidental changes or deletions. Also keep separate copies of every data transformation (version) made during your processing stages.

- **Backup your data and projects!** A useful mnemonic is the 3-2-1 rule: Keep 3 copies on 2 media types, with 1 remote copy. Media types can include local disk, cloud storage solutions, external disk, etc. Make regular backups, periodically test those backups to ensure you can recover data from them, and develop plans to backup software packages with version-specific features that may be critical to your work.

- **Use version control.** GitHub/GitLab/Bitbucket can help maintain a well-organized history of changes to your files. It's typically used for text-based files such as source code, scripts, etc. Several analytics platforms – including R Studio – integrate with GitHub.
  - You may self-provision a personal, publically accessible repository from GitHub at: https://github.com/personal
  - For additional security, you may prefer a repository from Engineering's instance of GitLab at: https://gitlab-beta.engr.illinois.edu/
  - You may provision a personal Bitbucket repository at: https://bitbucket.org/account/signup/
  - Box, Google, and Microsoft One Drive also provide some amount of version control.

- **Know your funders' requirements.** As of February of 2013, the federal government requires federally funded research be made public within a certain timeframe after publication and requiring that researchers should account for and manage the digital data created from federally funded scientific research. A list of funding agencies' publication requirements can be found here:
  http://www.library.illinois.edu/scholcomm/access_mandates/mandates.html