Week 2: File names, formats, & organization; Data Storage & Backup

updated: 2021-02-03

Pre-Class Preparation & Materials Needed (Instructor):

Send in an email to students:

- Confirmation of room and zoom link
- Remind students to bring their computers
- Make sure you know if everyone has R and instlled; level of R fluency
- Snacks
- Copies of the syllabus
- Copy of Course Roster
- Flip charts and markers
- Dry write markers
- Tent cards for student names

Objectives and Competencies for this session:

- Describe and implement conventions for proper naming of files
- Explain the difference between proprietary and open formats
- Learn how to efficiently organize their research data files
- Learn the preferred format for storing and archiving different types of data files
- Become familiar with different options for cloud data storage and backup
- Develop and implement a plan for short- and long-term data storage, back-up, and archiving
- Learn rules and policies for data security
- Become familiar with tools for such tasks as batch renaming of files, clud data storage, and automated data backup.
- Explain options for a long-term sustainable preservation strategy/policy for your data (e.g., discipline specific, institutional, departmental, individual).
- Address the need for conversion to standard formats needed for re-use
- Perform basic archival processes: checksum, auditing, format migration, etc.
- Understand costs & timelines for data storage, management tools and services

Pre-class Preparation (Students):

- **Readings:**

1. Jan Čurn. 2014. How a bug in Dropbox permanently deleted my 8000 photos. [read online] [download pdf]

2. PSA: Scrivener, Data Integrity and You. Or, How To Avoid Data Loss Heartbreak. [read online] [download pdf]

3. Hart EM et al. (2016) Ten Simple Rules for Digital Data Storage. PLoS Comput Biol 12(10): e1005097. [read online] [download pdf].

4. Panzarino, M. 2012. How Pixar's Toy Story 2 was deleted twice, once by technology and again for its own good. TNW. [read online] [[download pdf]]

- **Online Lectures:**

  1. Video on Project Organization
  2. Video on File Names
  3. Video on Storage and Backup

Class Outline

1. Turn on zoom captioning

2. **Block 1: (20 min):** Any questions form last week? Followed by an overview of today's activities.

3. **Block 2: File Format Jeopardy (35 min)**: These are common file formats used to collect, analyze, and store data. Can you tell me what the extensions stand for? Winner gets swag…

4. **Block 3: Breakout 1 (25+10 min):**

   1. Discussion of Data Sacurity and Backup in the field + snack break

5. **Block 4: Breakout 1 return results (30 min)**

6. **Block 5: Breakout 2 (60 min)** Interview Breakout, collect ansewrs on google doc or word doc and mail to me, remainder of time is to work on activity 2. Instructor regularly checking in on students

After class:

- Be sure you complete and submit the assignment by 9 am Monday
- Prepare for next session (assigned reading, videos, etc).

Alternative Questions:

1. How might you organize a time series of photos from several remote cameras? Possibility 1: images from each camera are in different folders. These are, in turn, collected in a folder named "images". Each folder would be named with the identifier for the camera. The file names for images might reflect the time the image was taken.