# Hands-on Activity 6: Data Protection & Backups

**Associated DataONE Lecture:** Lesson 6: *Data Protection and Backups*

**Objectives:** Students consider the steps that need to be taken to protect data, and multiple ways in which data could be lost if not protected.

**Outcomes:** Students can explain how data can be protected through various means of backing up, archiving, and preserving data.

**Time Needed:** In a classroom setting: one hour outside of class and one hour in class discussion. In a workshop setting: one hour.

**URLs:** http://data-archive.ac.uk/create-manage/storage

**Additional Files Needed**: None

**Key Reading:**  Veerle Van den Eynden, Louise Corti, Matthew Woollard, Libby Bishop and Laurence Horton. 2011. Managing and Sharing Data: Best Practices for Researchers. Third edition. Section: "Storing Your Data." Online: http://www.data-archive.ac.uk/media/2894/managingsharing.pdf.

## Notes and Instructions for Instructors:

In a *classroom* setting, students interview a researcher - their research advisor, a committee member, other researcher. Ask a series of scripted questions - file naming practices and backup/versioning, backups on and off site, archiving practices. Students share what they find in small group or as a class, discussing strengths & weaknesses of practices.

In a *workshop* setting, a paired activity allows trainees to interview each other about their research practices if everyone has some experience in research. If few people have a research background (e.g. computer scientists learning environmental data management), the workshop organizer may want to plan to bring in several researchers in person or by video chat for small group interviews.

## Student Instructions:

There are many unfortunate stories of disasters that wipe out years of scientific research – floods, fires, or just mistakes that lead to data loss – and even more stories of research losing a few hours or days of excellent work from a local system crash or other error. Data protection is a key element to consider in the data life cycle, and not too difficult to achieve if scientists make it part of their daily workflow.

Your job is to interview a practicing researcher about his/her data protection practices. In this conversation, you will both have an opportunity to explore some of the practices currently being used by researchers, some of the perceived obstacles to robust protection, and the opportunities that exist for making data protection more robust.

Some of the questions will focus on backups and archives, terms that are frequently used interchangeably but which are different. *Backups* are copies of original files made before the original is overwritten; backups are usually performed with high frequency, used to take periodic snapshots of data in case the current version is destroyed or lost. *Archive* refers to long-term preservation of the file, usually the final version or created at a major milestone in the project, used to preserve data for historical reference or potentially during disasters. Both are important components of data preservation, and it might be helpful to explain the difference in the terminology when you start your interview.

## Questions for your interview

Decide whether your interviewee would like to focus on his/her own individual research, or that of his/her research group; note which is being discussed. It's okay to skip some questions if you get into an especially rich topic or find that you have already answered it.

1. Are there formal policies that govern your data protection? (e.g. at a government lab or an explicit data management plan for a project or lab)
2. Who is responsible for data backups in your research?
3. What is the process that they undertake?
   a. How frequently?
   b. Partial (incremental) or full backup?
   c. Where are they backed up?
   d. What metadata accompanies these backups?
   e. How do you verify that a backup has been successfully performed?
   f. Can you read data off older backups? (e.g. file formats that are still accessible)
4. Who is responsible for data archiving in your research?
5. What is the process that they undertake for data archiving?
   a. Where are the data archived?
   b. What metadata accompany these backups?
   c. Who can access these data?
   d. Can you read data from your archives? (e.g. file formats that are still accessible)
6. Are there non-digital information sources that are critical for your work?
7. If so, how are they protected?
8. Do these practices fit your needs?

**Group discussion following interviews**

1. What are the common practices in data backup and archiving?

2. How might these practices be improved?

3. What do researchers seem to need in order to be more effective in data protection – e.g. access to repositories, training, etc.?

DataONE