

Introduction to Research Data Management

Version control

<http://hdl.handle.net/1969.1/164383>

Introduction

Version control is used to:

- Ensure what you do is transparent and reversible.
- Track your progress.
- Improve workflow reproducibility.
- Reduce work when reconstructing how and why changes were made.

Data stability

1. **Fixed datasets** never change after being collected or generated.
2. **Growing datasets** new data are added, but the old data are never changed or deleted.
3. **Revisable datasets** new data are added, and old data may be changed or deleted.

Implementing version control

Make copies of files as you work with them, and adjust the file names.

- Document changes separately in your data documentation.
- Note when and how you choose to delete old versions.

Implementing version control

Use a Version Control System.

- More investment at the beginning.
- Automatically tracks changes as you work.
- Progressively document changes within the system.

Simple: Saving files iteratively

- Recognizably name new versions according to a naming convention.
- Develop a general definition for what constitutes a new version.
- Log a description of the changes made between versions, and/or track the changes themselves.
- Discard obsolete versions.

Discussion

Why should you discard or delete obsolete versions of data files?

Robust: Using a Version Control System

Version Control Systems (VCS) are software tools to help you automate version control.

Subversion: Centralized Version Control System. Older. (<https://subversion.apache.org>)

Git: Distributed Version Control System. From software development community. Integrates with services like Github and Bitbucket (<https://git-scm.com/book/en/v2/Getting-Started-About-Version-Control>)

Mercurial: Distributed Version Control System. Simpler than Git. (<https://www.mercurial-scm.org>)

Conclusion

- Explored reason for using version control.
- Identified elements of version control.
- Introduced Version Control Systems.

References and resources

- DMPTool. "Data Management General Guidance" [website](https://dmptool.org/dm_guidance)
- Git. "Getting Started - About Version Control" [website](<https://git-scm.com/book/en/v2/Getting-Started-About-Version-Control>)