Introduction to Versioning

Version Database

file — Version 3
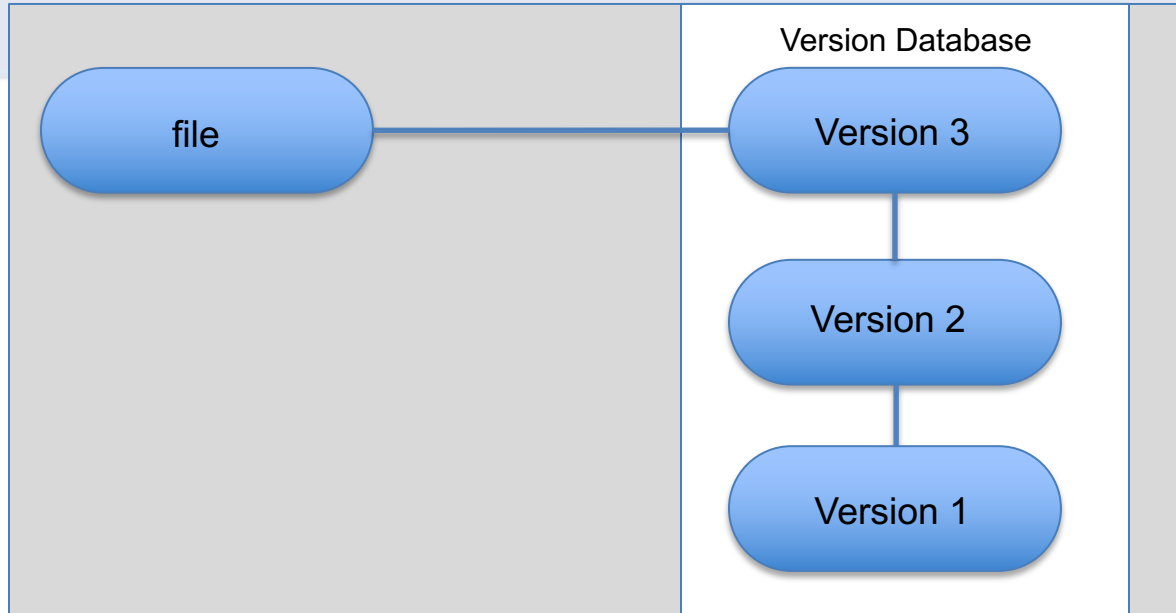
Version 2

Version 1

Sherry Lake

Data Management Consultant

University of Virginia Library

shLake@virginia.edu

Bill Corey

Data Management Consultant

University of Virginia Library

wtc2h@virginia.edu

Goals for the Workshop

- Identify version control problems and causes

- Understand differences between Version Control and Revision History

- Learn about *programs* that provide versioning

- Learn about **GitHub** an online version control repository

How do you…

- Identify the most recent version of a file?

- Go back to a previous state of the file?

- Tell the difference between versions of a file?

- Know which is the "real" final version?

- Keep from working on the same document at the same time?

What is Version Control?

Version control refers to file revisions management. It facilitates **best practice in research data management** during a project where constant redrafting and revision is occurring by numerous researchers.

Achieved by:

- Naming conventions

- File Revision (file history)

- Document Control Sheets

- Versioning Software

Naming Conventions

- Collaboration: few people

- Length of project: short

- Need training and documentations on how (use date/time/initials?)

- Does not prevent two people working on the "next" version (2 diff files) at the same time.

Naming Conventions

[yyyymmdd] [filename].[file extension]

Examples:

    20111218 Learning Spaces Program.doc

    20111211 Learning Spaces Program.doc

[filename]V[+1].[file extension]

Examples:

    Library-RenovationV5.doc

    Library-RenovationV4.doc

Make it easy to identify drafts and final versions at a glance

Revision History

Software that saves versions*:	
Wiki (i.e., UVa Collab)	Wordpress
UVa Box	MS Sharepoint
Google Drive	Evernote (premium feature $)

Features (not available on all listed above):
- Download previous versions
- Revert back to previous versions
- Compares 2 documents
- Step through drafts to visualize changes side-by-side

*List not inclusive

Document Control Sheets

- record details of the revision process
- who made the changes, when and why

Example document control sheet

File Name	Electronic Records Management PID d1b
Original Author(s)	Andrew Stewart
Current Revision Author(s)	Steve Bailey

Version	**Date**	**Author(s)**	**Notes on Revisions**
d1a	2006/01/12	AS	Initial start-up
d1b	2006/15/12	SB	More detail added to the initial PID

Data Provenance

- Document the origin and history (transformations) of a dataset

- Tools for documenting data file edits:
  - OpenRefine (formerly Google refine)
 - Statistical software packages
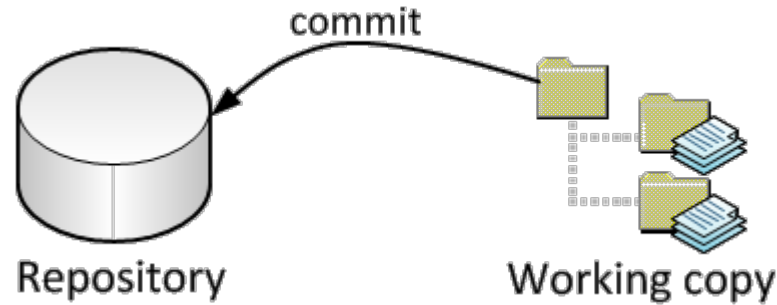
Version Control Systems

- Can function as additional backup

- Can go back to earlier version (or "last known good")

- Can document who & what changes

- Work on any type of document (not just software)

- Online service, free (be aware of privacy policies when using hosted services)
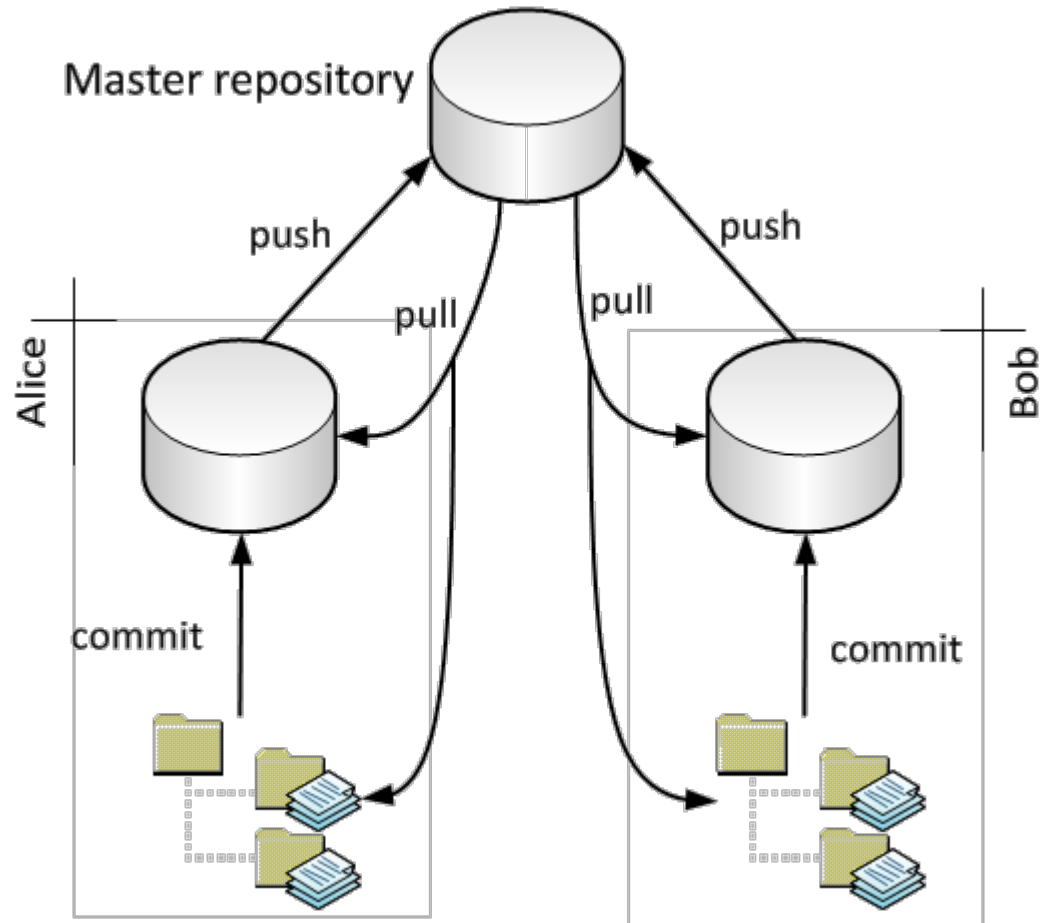
Version Control Systems

Basic concepts:

- Repository
- Server
- Client
- Add - Commit
- Check in / Check out
- Changelog / History
- Diff

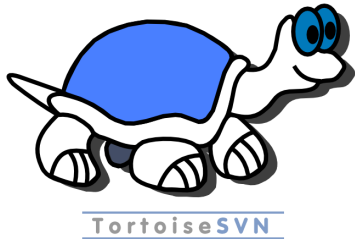- Advanced features: branching, merging, conflict resolution

UNIVERSITY
of VIRGINIA
LIBRARY

Version Control Graphical View

http://pages.cs.wisc.edu/~driscoll/software/vcs/
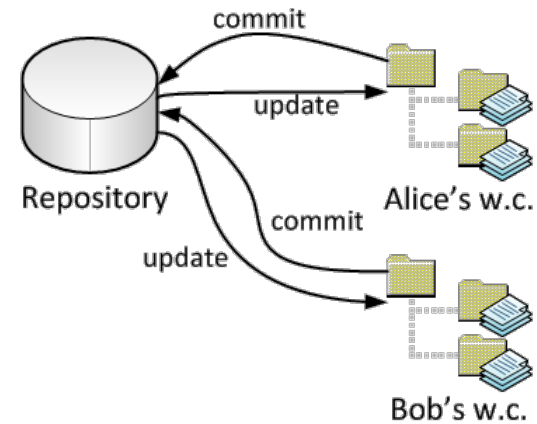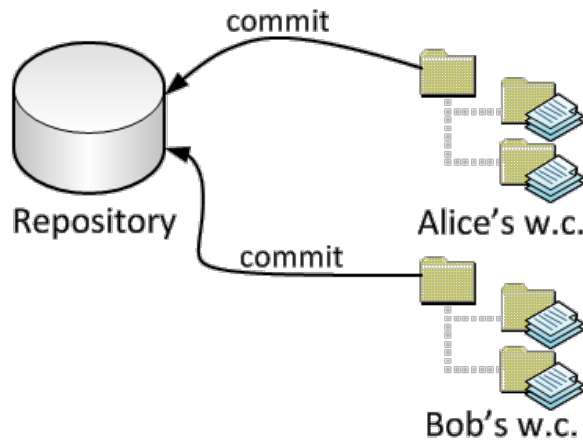
Distributed Version Control

http://pages.cs.wisc.edu/~driscoll/software/vcs/

Version Control Software

Examples:

Subversion (SVN) a local server
tortoiseSVN: windows client software

github.com: online repository (server)
git: client software

github Exercise

- Create github.com account

- Create a repository online

- Create a repository locally

- Add a file and Commit

More Information

- UVa Box
  http://its.virginia.edu/box/

- MS Sharepoint (UVa)
  http://its.virginia.edu/sharepoint/

- Github.com

- Github help
  https://help.github.com/articles/create-a-repo

RESEARCH DATA SERVICES

Offering expert data assistance at every stage of the research process.

1: PLANNING

We can assist you with developing a data management plan and designing your planned data analysis, including:

- Implementing plans, using tools, and creating workflows for managing research data
- Advising on study design, power analysis, and choice of statistical methods
- Helping to meet increasingly stringent criteria from funding agencies

2: FINDING & COLLECTING

We have access to thousands of sources of data and experts who will help you:

- Locate, evaluate and format data
- Create metadata and data documentation protocols for new data collection
- Capture data using best practices and appropriate technology

3: ANALYZING

Get expert assistance from statistical, spatial, or media specialists to analyze your data and present your research:

- Learn to use cutting-edge tools and methods
- Experiment with high-resolution visualization technologies
- Develop graphical representations that bring impact to your analysis

4: SHARING & ARCHIVING

We can consult with you on strategies to help others discover or access your research by:

- Adhering to data sharing policies and norms
- Selecting a data-sharing repository
- Making your data easier to discover and reuse

UNIVERSITY of VIRGINIA LIBRARY

Workshops • 1:1 Consultations • Class Presentations
Contact us at researchdataservices@virginia.edu