



Version Control

Preservation and Curation of ETD Research Data
and Complex Digital Objects



EDUCOPIA
INSTITUTE

Workshop Background

Purpose

- Provide an overview of the challenges and opportunities “version control” poses and provides for you as a researcher.

Context

- Workshop Series: Preservation and Curation of ETD Research Data and Complex Digital Objects
- Other topics: Copyright, Data Organization, File Formats, Metadata, Storage
- <https://educopia.org/research/etdplus>



Learning Objectives

- Understand that your research files will change over time
- Learn to use naming/storage conventions to “version” your research files, making it easier to know your research history
- Gain familiarity with software tools that can help you with versioning, particularly when you’re collaborating on documents with others

Version Control

- The process of managing changes to your files over time (aka, revision control or source control).

Version 01

Version 02

Version 03

Version 04

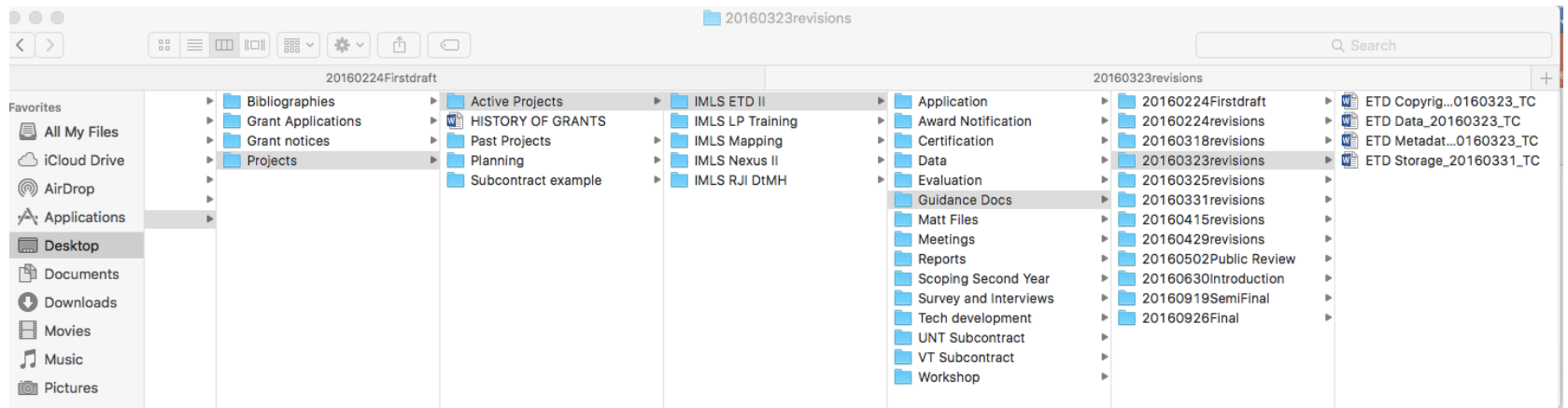
Version 05

Why Versions Happen

- Saving a new draft of text for editing
- Refining raw survey results into a clean dataset for analysis
- Producing a transcript of an interview based on an audio recording
- Creating a smaller version of an image to post online

Photo by Rene Bohmer on Unsplash

Version Control is all about PROCESS



Version Control

OK

image1_v1.jpg
image1_v2.jpg
image2_v1.jpg
image2_v2.jpg
...

Oops

image1_v1.jpg
image1_v10.jpg
image1_v2.jpg
...

Better

image1_20151021
image1_20151214
image1_20160123
...

Version Control – Collaborative Docs

- dataset1_20160402_KES
- dataset1_20160301_WTC
- dataset1_20160814_GSC
- ...
- ...

Versioning Tools

- Examples: Github and Subversion
- Check-in and check-out file processes
- Usually you can only view and edit the working version of a file.
- Each time you change a file, you can save a “revision” and attach a short summary of your changes.

Resources

- MATRIX at Michigan State University gives file naming advice: <http://ohda.matrix.msu.edu/2012/08/file-naming-in-the-digital-age>
- Udacity offers a free online course on using Git and GitHub: <https://www.udacity.com/course/how-to-use-git-and-github--ud775>
- Hello World offers another helpful GitHub guide: <https://guides.github.com/activities/hello-world/>
- *Version Control with Subversion* is a free book authored by Subversion software developers: <http://svnbook.red-bean.com/>

Version Control

Version Control: The process of managing changes to your files over time (aka, revision control or source control)

Manual Version Control

A simple method to store the current revision is at the end of the file name. This way, files can be grouped by their names and sorted by version number:

- filename-v01.jpg
- filename-v02.jpg
- ...

You can also use dates to designate version numbers, using year-month-day (20150930) to help your computer sort versions in chronological order:

- filename-20160402.jpg
- filename-20160407.jpg
- ...

If the files you are using are created or edited collaboratively, incorporate names or initials so you know who updated which version:

- filename-20160402-KES.jpg
- filename-20160407-WTC.jpg
- ...

Software-Assisted Version Control

There are also software tools that can help you version your content. These tools store your content in such a way that they can remember its state from revision to revision. Usually, they also allow you to “check in” and “check out” your content, ensuring that revisions never happen simultaneously in two different locations (e.g., if collaborating researchers both attempt to revise the same file at the same time, or a researcher unwittingly tries to revise the same file on two different machines). Key differences between these software-assisted methods and the manual methods include:

1. You can only view and edit the working version of a file
2. When you change a file, you can save a revision and attach a short summary of your changes.

Research is active and iterative. You will edit and re-edit your research materials many times before finishing your thesis or dissertation. How will you know that you are working with the most current revision of your materials?

Resources (For more information)

- The digital humanities center MATRIX (Michigan State University) provides advice on how to structure file names based on oral history projects that is broadly applicable: <http://ohda.matrix.msu.edu/2012/08/file-naming-in-the-digital-age>
- Udacity offers a free online course on how to use Git and GitHub with interactive exercises to familiarize you with using the tools. <https://www.udacity.com/course/how-to-use-git-and-github--ud775>
- Another helpful GitHub guide is available from Hello World. <https://guides.github.com/activities/hello-world/>
- The Subversion community provides free access to the book Version Control with Subversion: <http://svnbook.red-bean.com/>



Activity

- Find a folder of research materials that you have collected on your computer. Look through the materials and answer the following questions:
 - Are there multiple versions of the same materials (documents, images, etc.)?
 - How are the different versions labeled?
 - Can you quickly identify a file's most recent version? Its authoritative version? Its original version?

