

Workshop: High-performance computing for economists

Lars Vilhuber¹ John M. Abowd¹ Richard Mansfield¹
Kevin L. McKinney

¹Cornell University, Economics Department,

August 20-22, 2013: Day 2

What are VCS

Software development perspective

Version control systems (VCS) (or Source Configuration Management (SCM) systems *) allow developers or authors to keep track of the history of their projects? source code. [source]

What are VCS

Software development perspective

Version control systems (VCS) (or Source Configuration Management (SCM) systems *) allow developers or authors to keep track of the history of their projects? source code. [source]

Generic view

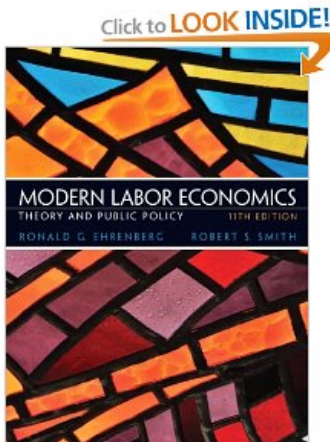
Detailed mechanism to manage different versions (historical, parallel) of documents, files, programs, etc.

You already have it

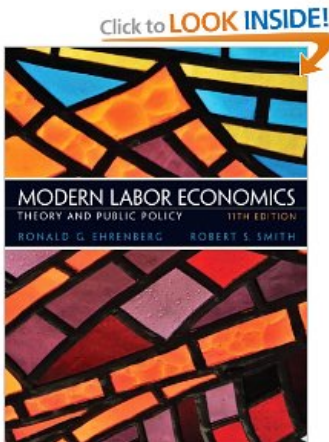
Implicit uses

- ▶ Backup systems (Apple Time Machine, others)
- ▶ Word processors (Undo feature, Track changes in Word, finer-grained features in content management systems: Google Docs, Blog software, etc.)
- ▶ Versioning filesystems
- ▶ Paper books!

Modern Labor Economics: Theory and Public Policy



Modern Labor Economics: Theory and Public Policy



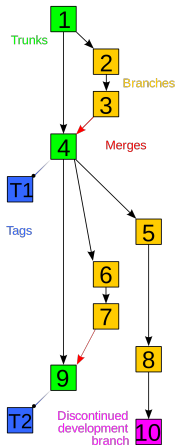
Preface

New to This Edition

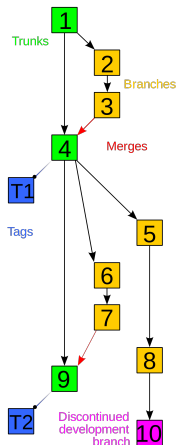
- This eleventh edition of *Modern Labor Economics* has been thoroughly updated in terms of both tabular material and references to the latest literature. Our goal in these updates is to make our textbook a comprehensive reference, for both students and professors, to critical factual information about the labor market *and* to the professional literature in labor economics.
- In recognition of the growing need for rigorous and dispassionate analyses of American immigration policy, we have expanded our analysis of undocumented immigration in [chapter 10](#) to include an enhanced analysis of both its theoretical and measured effects on society.
- We have also incorporated, in relevant chapters, discussions that include labor-market effects of the Great Recession of 2008, along with an examination of recent changes in such outcomes as earnings inequality, human-capital acquisition, and labor-force participation.
- In [chapter 6](#), we added a discussion of the labor supply behavior of married women and a new boxed example on the labor supply of New York City taxi drivers.
- In [chapter 11](#), we amplified the “Group Incentives and Executive Pay” section and added a new boxed example on the “rat race” in law firms.
- In addition to including new material on the recession, we added a new boxed example on earnings inequality in developed countries and a new section on earnings instability to [chapter 15](#).

Principal idea

1 First edition



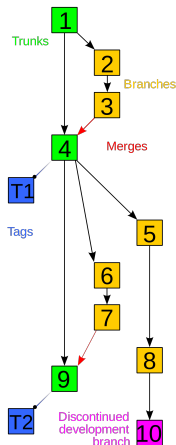
Principal idea



1 First edition

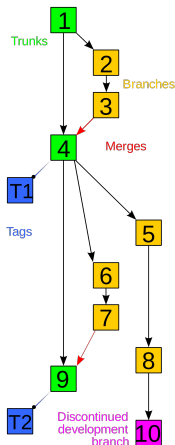
4 Second edition

Principal idea



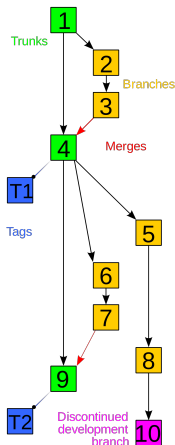
- 1 First edition
- 4 Second edition
- 5 Start of work on a Canadian edition

Principal idea



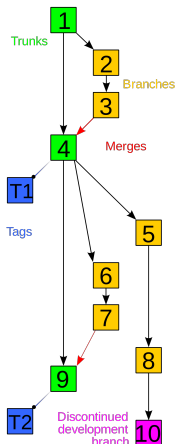
- 1 First edition
- 4 Second edition
- 5 Start of work on a Canadian edition
- 6 Start of work on the next US edition

Principal idea



- 1 First edition
- 4 Second edition
- 5 Start of work on a Canadian edition
- 6 Start of work on the next US edition
- 9 Third edition

Principal idea



- 1 First edition
- 4 Second edition
- 5 Start of work on a Canadian edition
- 6 Start of work on the next US edition
- 9 Third edition
- 10 First Canadian edition

Two major types of version-management

Centralized model

Server-client approach, editors check out a copy, modify it, and check it back in. Multiple editors:

- ▶ *File locking*: only one person can check out any given file
- ▶ *Version merging*: discrepancies are handled upon checkin

Two major types of version-management

Centralized model

Server-client approach, editors check out a copy, modify it, and check it back in. Multiple editors:

- ▶ *File locking*: only one person can check out any given file
- ▶ *Version merging*: discrepancies are handled upon checkin

Distributed model

There is no central server (prescribed by software). Every editor has a full copy of all version, synchronisation occurs by exchanging patches.

Focus in this class

We will focus on one system primarily, and mention a second one:

- ▶ Subversion (centralized)

Windows TortoiseSVN (free)

OSX Versions (\$), Xcode (free)

- ▶ Git (decentralized)

Quickstart on SSG/ECCO

The following really quick tutorial can be run on SSG or on your own computer, if you already have the software installed.

```
ssh NETID@ssg.vrdc.cornell.edu
```

```
Last login: Mon Aug 19 15:43:33 2013 from lv39-ws.ilr.cornell.edu
```

```
Welcome to the Social Science Gateway (SSG),
for help, send email to ssg-help@cac.cornell.edu
```

```
ssg:~> svn help
```

```
usage: svn <subcommand> [options] [args]
```

```
Subversion command-line client, version 1.6.6.
```

```
Type 'svn help <subcommand>' for help on a specific subcommand.
```

```
Type 'svn --version' to see the program version and RA modules
```

```
    or 'svn --version --quiet' to see just the version number.
```

Most subcommands take file and/or directory arguments, recursing on the directories. If no arguments are supplied to such a **command**, it recurses on the current directory (inclusive) by default.

Available subcommands:

add

blame (praise, annotate, ann)

cat

[snip]

Quickstart Step 1

Let's create a place to work

```
mkdir Workspace  
cd Workspace
```

Quickstart Step 1

Let's create a place to work

```
mkdir Workspace  
cd Workspace
```

Check out a version of the code:

```
> svn co http://repository.vrdc.cornell.edu/public/test  
A   test/test.3  
A   test/test.4  
A   test/releases  
A   test/trunk  
A   test/trunk/bls  
[snip]  
A   test/test.1  
Checked out revision 23.  
>
```

Checkout To *check out* (or *co*) is to create a local working copy from the repository. A user may specify a specific revision or obtain the latest. The term 'checkout' can also be used as a noun to describe the working copy.

Quickstart 2

Using your favorite editor, modify a file:

```
> cd test  
> vi test.1
```

Quickstart 2

Using your favorite editor, modify a file:

```
> cd test  
> vi test.1
```

Verify the state of the repository

```
> svn status  
M      test.1
```

Quickstart 2

Using your favorite editor, modify a file:

```
> cd test
> vi test.1
```

Verify the state of the repository

```
> svn status
M      test.1
```

Let's also add a file:

```
> vi test.today
> svn status
?      test.today
M      test.1
```

The question mark indicates that the file is not (yet) versioned.

Quickstart 3

Add the unversioned file to Subversion tracking

```
> svn add test.today
A      test.today
> svn status
M      test.1
A      test.today
```

Quickstart 3

Add the unversioned file to Subversion tracking

```
> svn add test.today
A      test.today
> svn status
M      test.1
A      test.today
```

Now we can commit our changes

```
> svn commit
[editor pops up]
Sending      test.1
Adding      test.today
Transmitting file data ..
Committed revision 24.
```

Commit To *commit* (*check in*, *ci*) is to write or merge the changes made in the working copy back to the repository. The terms 'commit' and 'checkin' can also be used as nouns to describe the new revision that is created as a result of committing.

Additional things you will want to learn

Manipulating files

- ▶ `svn mv` (vs regular Linux `mv` or cut-and-paste)
- ▶ Reverting changes (`svn revert` before a commit, more complicated after a commit)
- ▶ Merging changes (`svn merge`)
- ▶ Inspecting logs (`svn log`)
- ▶ Identifying changes relative to previous revisions (`svn diff`)

THE reference for Subversion

The Standard in Open Source Version Control

Version Control with Subversion

Introduction

This is the home of *Version Control with Subversion*, a [free](#) book about the ubiquitous [Apache™ Subversion](#) version control system and written by some of the developers of Subversion itself.

As you may have guessed from the layout of this page, we are quite pleased that some versions of this book have been [published](#) by O'Reilly Media. You can certainly [buy](#) a traditionally published print copy of the book if you'd like to, but you'll always find the most recent versions here on this site, available in both HTML and PDF formats.

Online Versions of the Book

Versions of this book use a numbering system designed to match those used by the Subversion software itself—version 1.7 of *Version Control with Subversion* covers Subversion 1.7, for example. Here are the latest versions of the book which are available online:

For Subversion 1.7

- View the [multiple-page HTML edition](#) of the book. [[tar.bz](#)]
- View the [single-page HTML edition](#) of the book. [[tar.bz](#)]
- View the [PDF edition](#) of the book.
- View the book's [DocBook XML sources](#).

Nightly Build (for Subversion 1.8)

Please bear in mind that these versions are works-in-progress if you bookmark or link to specific sections; those links may be invalidated by continuing development. If you need a link that can be reasonably expected to remain stable for years to come, link to one of the completed editions above.

- View the [multiple-page HTML edition](#) of the book. [[tar.bz](#)]
- View the book's [DocBook XML sources](#).

svnbook.red-bean.com

Learning more

A longer tutorial

A longer tutorial (focussing on a complex project setup) is available at the course website.

Tracking history

What happened? And who did what?

One of the key advantages of using version control systems is ... to control versions.

- ▶ Straightforward to view multiple versions of a file (assuming proper usage)
- ▶ Possibility to view who changed what (“blame” or “annotate”)

Tracking using web interfaces: SVN

LEHD Code Repository

logged in as vilhuber | Logout | Help/Guide | Preferences | About Trac

Subversion Location | Last Change | Revision Log

root / trunk

Visit: View revision:

Name ▲	Size	Rev	Age	Last Change
..				
admin		8646	7 months	beau001: Copy physical parms file during refresh dev. Re. #957
bls		9187	4 months	beau001: Move macros to global process. Update Trunk. Re #995
brb		1492	7 years	vilhuber: Re-establishing fully tested trunk as per production 2006-10-03 and prior
control		9727	3 weeks	beau001: Updating Trunk. Modify code_copy and check_versions to read the correct ...
cpr		4524	3 years	vilhuber: Re #461: this switches the preprocess parameter to preprocess=yes, as per ...
doc		9368	4 months	beau001: This is the Quality Audit Template that has all of the standards that the ...
ecf		9712	3 weeks	beau001: Updating trunk. Read new format for GALID. Re. #997 #1001 #1022 #1033 ...
edf		8101	10 months	beau001: Cleanup process directory. Re. #898
ehf		8102	10 months	beau001: Cleanup process directory. Re. #898
ctrlprogs		7220	17 months	beau001: Re #272, #593, #781. Merge into trunk changes for getting EIN info from ...
control_ehf.sas	17.4 KB	7220	17 months	beau001: Re #272, #593, #781. Merge into trunk changes for getting EIN info from ...
parameters_ehf.sas	1.8 KB	7220	17 months	beau001: Re #272, #593, #781. Merge into trunk changes for getting EIN info from ...
doc		8102	10 months	beau001: Cleanup process directory. Re. #898
library		7649	14 months	beau001: Change methodology for creating ehf_<state>_ui_ein. Re: #870.
macros		7649	14 months	beau001: Change methodology for creating ehf_<state>_ui_ein. Re: #870.
sasprogs		7220	17 months	beau001: Re #272, #593, #781. Merge into trunk changes for getting EIN info from ...
01_ehf.sas	2.5 KB	7220	17 months	beau001: Re #272, #593, #781. Merge into trunk changes for getting EIN info from ...
01_ehfqa.sas	7.9 KB	4028	3 years	vilhuber: Re #629. Based on ticket-629/ehf (r3067:3910). New QA test by Stephen
02_ehf.sas	3.2 KB	7220	17 months	beau001: Re #272, #593, #781. Merge into trunk changes for getting EIN info from ...
02_ehfqa.sas	2.7 KB	4028	3 years	vilhuber: Re #629. Based on ticket-629/ehf (r3067:3910). New QA test by Stephen
03_ehf.sas	1.7 KB	687	8 years	vilhuber: Legacy change EHF 3.1.30 Time-stamp: <04/08/26 16:43:07 vilhu001> * added ...
03_ehfqa.sas	0.7 KB	4028	3 years	vilhuber: Re #629. Based on ticket-629/ehf (r3067:3910). New QA test by Stephen
04_ehf.sas	1.7 KB	687	8 years	vilhuber: Legacy change EHF 3.1.30 Time-stamp: <04/08/26 16:43:07 vilhu001> * added ...
04_ehfqa.sas	1.1 KB	4028	3 years	vilhuber: Re #629. Based on ticket-629/ehf (r3067:3910). New QA test by Stephen
05_ehf.sas	5.3 KB	7220	17 months	beau001: Re #272, #593, #781. Merge into trunk changes for getting EIN info from ...
05_ehfqa.sas	0.7 KB	4028	3 years	vilhuber: Re #629. Based on ticket-629/ehf (r3067:3910). New QA test by Stephen
06_ehf.sas	2.9 KB	7220	17 months	beau001: Re #272, #593, #781. Merge into trunk changes for getting EIN info from ...

enhancem FIRST_RUN

Vilhuber, Abowd, Mansfield, McKinney

Computing for Economists

Tracking using web interfaces: SVN

LEHD Code Repository

logged in as vilhuber | [Logout](#) | [Help/Guide](#) | [Preferences](#) | [About Trac](#)

[Wiki](#) | [Browse Source](#) | [View Tickets](#) | [New Ticket](#) | [Timeline](#) | [Roadmap](#) | [Rev Tree](#) | [Search](#) | [Admin](#)

View Latest Revision

root / trunk / ehf / library / sasprogs / 01_ehf.sas

Added
Modified
Copied or renamed

View changes

Revision Log Mode:

- ☐ Stop on copy
- ☒ Follow copies
- ☐ Show only adds and deletes

View log starting at and back to
Show at most revisions per page.
☐ Show full log messages

Update

Diff	Rev	Chgset	Date	Author	Log Message
	@7220 [7220]	17 months	beauz001	Re #272 , #593 , #761 . Merge into trunk changes for getting EIN info from ...	
	@695 [695]	8 years	lv39	Moving EHF to trunk	
copied from merge_ehf/trunk/ehf/library/sasprogs/01_ehf.sas:					
	@676 [676]	8 years	vilhuber	Legacy change EHF 3.1.26-lars	
	@670 [670]	8 years	vilhuber	Legacy change EHF 3.1.23	
	@644 [644]	8 years	vilhuber	Legacy change EHF 3.1.10	
	@640 [640]	8 years	vilhuber	Legacy change EHF 3.1.08	
	@635 [635]	8 years	vilhuber	Legacy change EHF 3.1.05	
	@629 [629]	8 years	vilhuber	Legacy change EHF 3.1.02	
	@627 [627]	8 years	vilhuber	This is a legacy commit of ehf 3.1.01	

Note: See [TracRevisionLog](#) for help on using the revision log.

Download in other formats:
 RSS Feed | [ChangeLog](#)

Tracking using web interfaces: SVN

Update

☐ Unmodified ☒ Added ☐ Removed ☐ Modified

trunk/qwi/library/sasprogs/01_qwi.sas

Tabular Unified

• Property svn:keywords set to Id URL

Revision 552	Revision 9714
1 /* +--< LEHD-QWI qwi 3.1.10 2003-11-10 schwa305 >-- */	1 /* \$Id\$ */
2 /* +--< Location: /programs/production/current/qwi/ >-- */	2 /* \$URL\$ */
3 /* +--< File: library/sasprogs/01_qwi.sas >-- */	3 /* Author:*/
4	4 /* Time-stamp: */
5 *****	5 *****
6 * Time-stamp: <03/06/08 01:11:28 steph305>	6 * Time-stamp: <04/06/17 12:28:14 vilhuber>
7 * Author: Simon Woodcock	7 * Author: Simon Woodcock
8 * Based on part of Jobflow01 by John Abowd, Karen Conneely,	8 * Based on part of Jobflow01 by John Abowd, Karen Conneely,
9 * Robert Pedace, and Paul Lengersmann	9 * Robert Pedace, and Paul Lengersmann
10	10
11 * Initial processing of individual and employment history master files	11 * Initial processing of individual and employment history master files
12	12
13 * QWI STEP 1: BUILD THE PERSON HISTORY FILE	13 * QWI STEP 1: BUILD THE PERSON HISTORY FILE
14	14
15	15
16	16
17 * Create soft link to this file and MACROLIST.SAS in state-specific	17
18 * directory, create CONFIG_ST.SAS in same directory, and run.	18
19	19
20 * REQUIRES-SAS: macrolist.sas	20
21 * REQUIRES-SAS: config_st.sas	21
22	22
23 * REQUIRES-DATA: ehf_[st].sas7bdat (all states but MN)	23 * REQUIRES-DATA: ehf_[st].sas7bdat (all states but MN)
24 * REQUIRES-DATA: ehf_seinunit_an.sas7bdat (MN only)	24 * REQUIRES-DATA: ehf_seinunit_an.sas7bdat (MN only)
25 * REQUIRES-DATA: [st]_u2wi.sas7bdat	25 * REQUIRES-DATA: [st]_u2wi.sas7bdat
26	26
27 * PROVIDES-DATA: phf_[st]_a.sas7bdat	27 * PROVIDES-DATA: phf_[st]_a.sas7bdat
28	28
29	29
30	30
31	31
32 * REVISION HISTORY:	32 * REVISION HISTORY:
33 * 2003-05-19: options statement moved to CONFIG_ST.SAS, comments updated by Simon	33 * 2003-05-19: options statement moved to CONFIG_ST.SAS, comments updated by Simon
34 * 2003-05-30: Bryce added x command to make soft links to output files	34 * 2003-05-30: Bryce added x command to make soft links to output files
35 * 2003-06-07: Bryce modified filling of earnings array to include only values in the valid QWI processing range	35 * 2003-06-07: Bryce modified filling of earnings array to include only values in the valid QWI processing range

Vilhuber, Abowd, Mansfield, McKinney

Computing for Economists

Tracking using web interfaces: SVN

LEHD Code Repository

[logged in as vilhuber](#)
[Logout](#)
[Help/Guide](#)
[Preferences](#)
[About Trac](#)

[Wiki](#)
[Browse Source](#)
[View Tickets](#)
[New Ticket](#)
[Timeline](#)
[Roadmap](#)
[Rev Tree](#)
[Search](#)
[Admin](#)

[Subversion Location](#)
[Last Change](#)
[Normal](#)
[Revision Log](#)

[root](#) / [trunk](#) / [qwi](#) / [library](#) / [sasprogs](#) / [01_qwi.sas](#) @ 9714

Visit:
View revision:

Revision 9714, 1.8 KB (checked in by beauz001, 3 weeks ago)

Updating Trunk. Re: [#433](#) [#554](#) [#1003](#) [#1010](#) [#1016](#) [#1020](#) [#1021](#) [#1026](#) [#1031](#) [#1032](#) [#1044](#) [#1054](#) [#1056](#)

Property svn:executable set to *

Property svn:keywords set to Id URL

Rev	Line
[7204]	1
[6733]	2
	3
	4
[558]	5
[976]	6
[998]	7
	8
	9
	10
	11
	12
	13
	14
	15
	16
	17
	18
	19
	20
	21
	22
	23
	24
	25
	26
	27
	28
[9714]	29
[558]	30
	31
	32
	33
	34
	35

```

/* $Id$ */
/* $URL$ */
/* Author:*/
/* Time-stamp: */
*****
* Time-stamp: <04/05/17 12:28:14 vilhuber>
* Author: Simon Woodcock
* Based on part of Jobflow01 by John Abowd, Karen Conneely,
* Robert Pedace, and Paul Lengermann
*
* Initial processing of individual and employment history master files
*
* QWI STEP 1: BUILD THE PERSON HISTORY FILE
*
*
* REQUIRES-DATA: ehf_[st].sas7bdat (all states but MN)
* REQUIRES-DATA: ehf_seinunit_mn.sas7bdat (MN only)
* REQUIRES-DATA: [st]_u2wi.sas7bdat
*
* PROVIDES-DATA: phf_[st]_a.sas7bdat
*
* .....;
* REVISION HISTORY:
*
* 2003-05-19: options statement moved to CONFIG.ST.SAS, comments updated by Simon
* 2003-05-30: Bryce added x command to make soft links to output files
* 2003-06-07: Bryce modified filling of earnings array to include only values in the valid QWI processing range
* 2013-05-23: Cleanup commented out code. (Added comment for Stephen Tibbets change.)
*****;
*****;
title1 "Create the Person History File (PHF)";
title2 "/programs/projects/EDE/v3.0/6st/.01.create_phf.sas";


```

Tracking using web interfaces: SVN

Including our demo change
VRDC repository viewer

Tracking using web interfaces: git

Bitbucket Repositories Create ?

 **Computing for Economists**
 computing4economists Share

Clone Fork Compare Pull request

Overview **Source** Commits Pull requests Downloads 1

master Computing for Economists / web / mk_pgms.sh Source Diff History

baa15b6 5 minutes ago Full commit Blame Raw Edit

```

1  #!/bin/bash
2  outfile=programs.md
3  cat > $outfile << EOF
4  High-performance computing for Economists - Programs
5  =====
6  EOF
7
8  function add_day {
9  cat >> $outfile << EOF
10
11  Day $1
12  ----
13  EOF
14  for file in $(find ../programs/day$1 -name \*sh -o -name \*R -o -name \*sas| sort)
15  do
16  echo "+ [{file##*/}]({file})" >> $outfile
17  done
18  }
19
20  add_day 1
21  add_day 2
22  add_day 3
23
24  multimarkdown $outfile > ${basename $outfile .md}.html
25  git add $outfile ${basename $outfile .md}.html
  
```

Tracking using web interfaces: git

Bitbucket Repositories [Create](#) ?

Computing for Economists computing4economists Share

[Clone](#) [Fork](#) [Compare](#) [Pull request](#)

[Overview](#) **[Source](#)** [Commits](#) [Pull requests](#) [Downloads](#) 1

master [Computing for Economists / web / mk_pgms.sh](#) [Source](#) [Diff](#) [History](#)

Author	Revision	Message	Date
Lars Vilhuber	baa15b6	Added a function in bash	6 minutes ago
Lars Vilhuber	c581940	Switching the tex files to use package lstlisting, adding separate directory for downloadable	22 minutes ago

Tracking using web interfaces: git

master | Computing for Economists / web / mk_pgms.sh

Source Diff History

Diff from c581940 22 minutes ago to baa15b6 6 minutes ago

web/mk_pgms.sh Side-by-side diff

```

3 3  cat > $outfile << EOF
4 4  High-performance computing for Economists - Programs
5 5  =====
6 6  +EOF
7 7  +
8 8  +function add_day {
9 9  +cat >> $outfile << EOF
6 10
7 10  -Day 1
11 11  +Day $1
8 12  - - - -
9 13  EOF
10 13  -
11 13  -for file in $(find ../programs/day1 -name \*sh -o -name \*R -o -name \*sas| sort)
14 14  +for file in $(find ../programs/day$1 -name \*sh -o -name \*R -o -name \*sas| sort)
12 15  do
13 16     echo "* [${file##*/}](${file})" >> $outfile
14 17  done
18 17  +}
19 17  +
20 20  +add_day 1
21 21  +add_day 2
22 22  +add_day 3
15 23
16 24  multimarkdown $outfile > $(basename $outfile .md).html
17 25  git add $outfile $(basename $outfile .md).html

```

Subversion vs. Git

incomplete

Checking out a subdirectory

SVN

```
svn co ${URL}/sub/directory (name)
```

GIT

```
mkdir <repo> && cd <repo>
git init
git remote add -f <name> <url>
git config core.sparsecheckout true
echo some/dir/ >> .git/info/sparse-checkout
echo another/sub/tree >>
.git/info/sparse-checkout
git pull <remote> <branch>
```

Source: [here](#)

VCS infrastructure

At Cornell

Cornell Sourceforge (see instructions on how to create)

Elsewhere

- ▶ CloudForge (free for personal use)
- ▶ GitHub (Git, subversion, free for open-source)
- ▶ BitBucket (Git, no subversion, free for academic users)
- ▶ Roll your own: very simple webserver setup for Subversion, no server required for Git on your PC

Next section