Git

a tale told by an idiot
Matt Burton (@mcburton)

Git is a *distributed version control system* (DVCS) or *source code management* (SCM) system

"I'm an egotistical bastard, and I name all my projects after myself. First 'Linux', now 'git'." - Linus Torvalds

push it
pull it
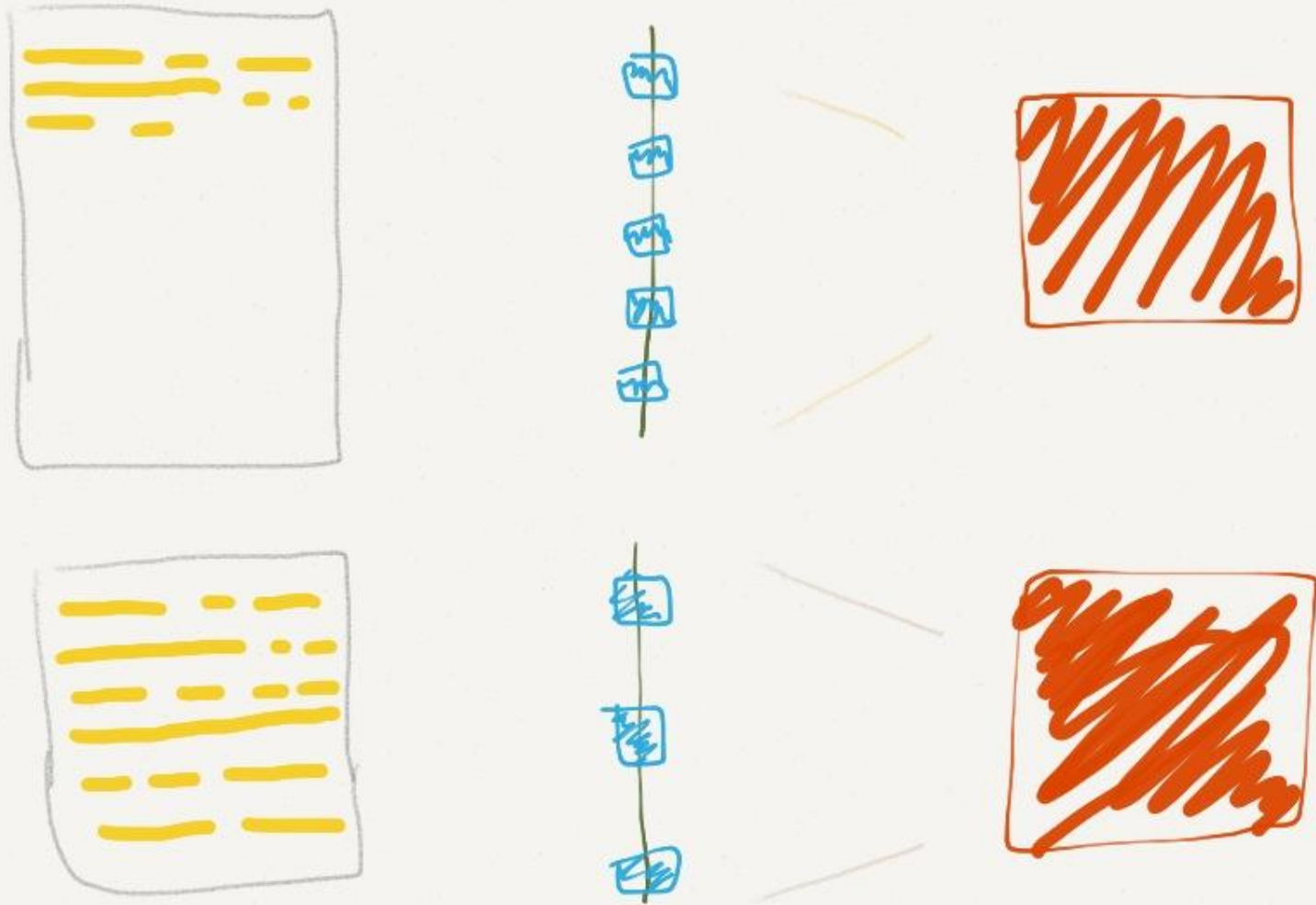**fork** it
clone it
check it
branch it
stash rebase it.

commit
merge it
log it
help it
clean it
blame it
move remove it

add it
grep it
diff it
fetch it
tag it
show it
pull request it.

Git's (complex) logic

Work you have saved can be bundled into a `Commit.`

A **Commit** is a snapshot of all *new* work since the previous **Commit**.

Commits are *humane* units of work. All glory to the metadata.

May 22, 2012

adding the ACH jobs mysql dump, sqlite3 database & a shell script to ... [– ...]
| ...convert mysql dumps to sqlite3 db.

mcburton authored 16 days ago

`9ec620286d` ⊕
Browse code ➡

May 16, 2012

Adding a pile of jobs
mcburton authored 22 days ago

`30e3827ea6` ⊕
Browse code ➡

May 01, 2012

adding glasgow
mcburton authored a month ago

`986c3a0371` ⊕
Browse code ➡

adding a few jobs & template
mcburton authored a month ago

`3e86143bde` ⊕
Browse code ➡

Apr 30, 2012

ok ya no
mcburton authored a month ago

`4b636c51a0` ⊕
Browse code ➡

commenting out responsive css for a test
mcburton authored a month ago

`aa0634ba2f` ⊕
Browse code ➡

adding headings
mcburton authored a month ago
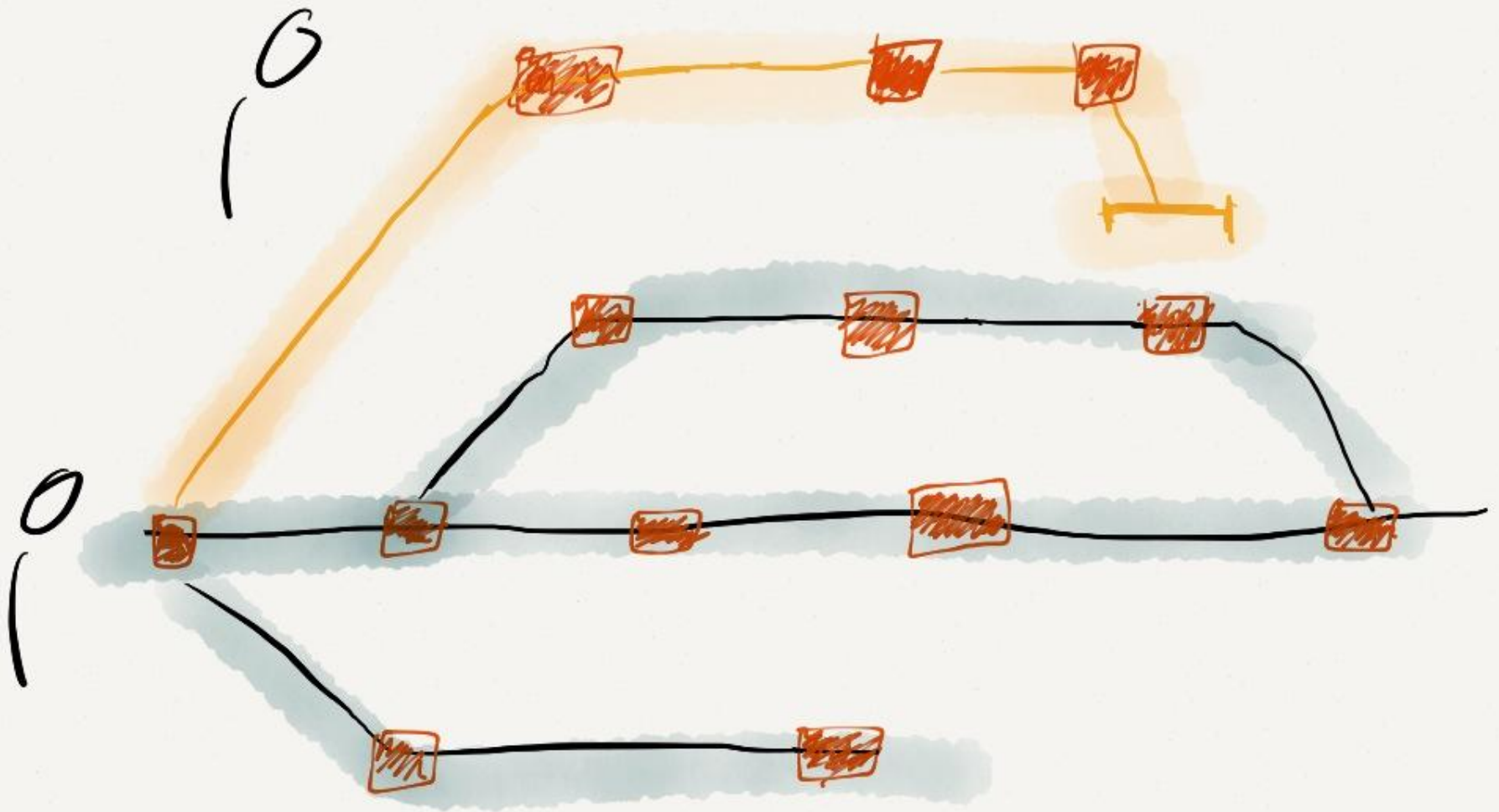
`af0877af25` ⊕
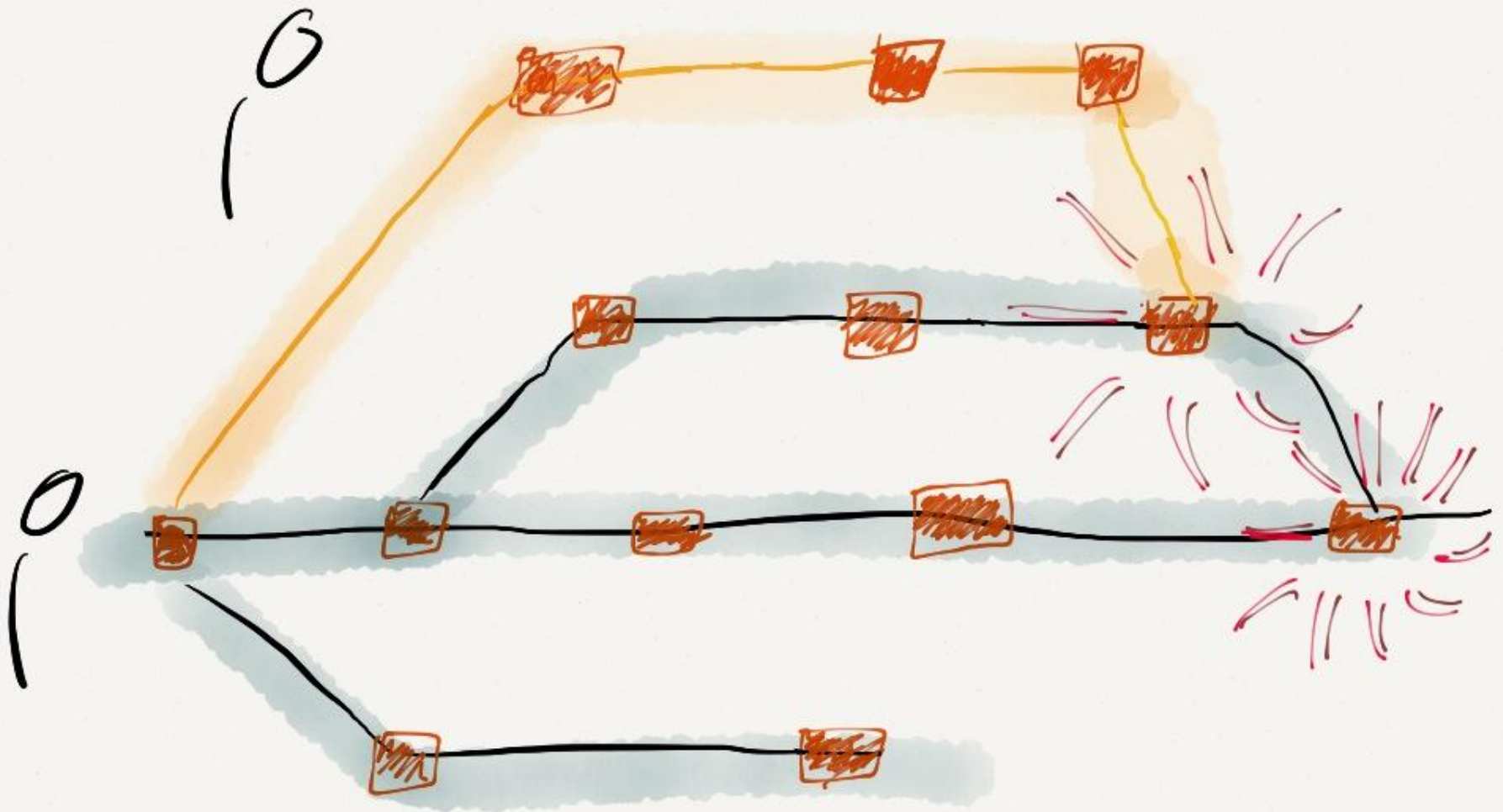Browse code ➡

fixing link

`1728bccaaa` ⊕

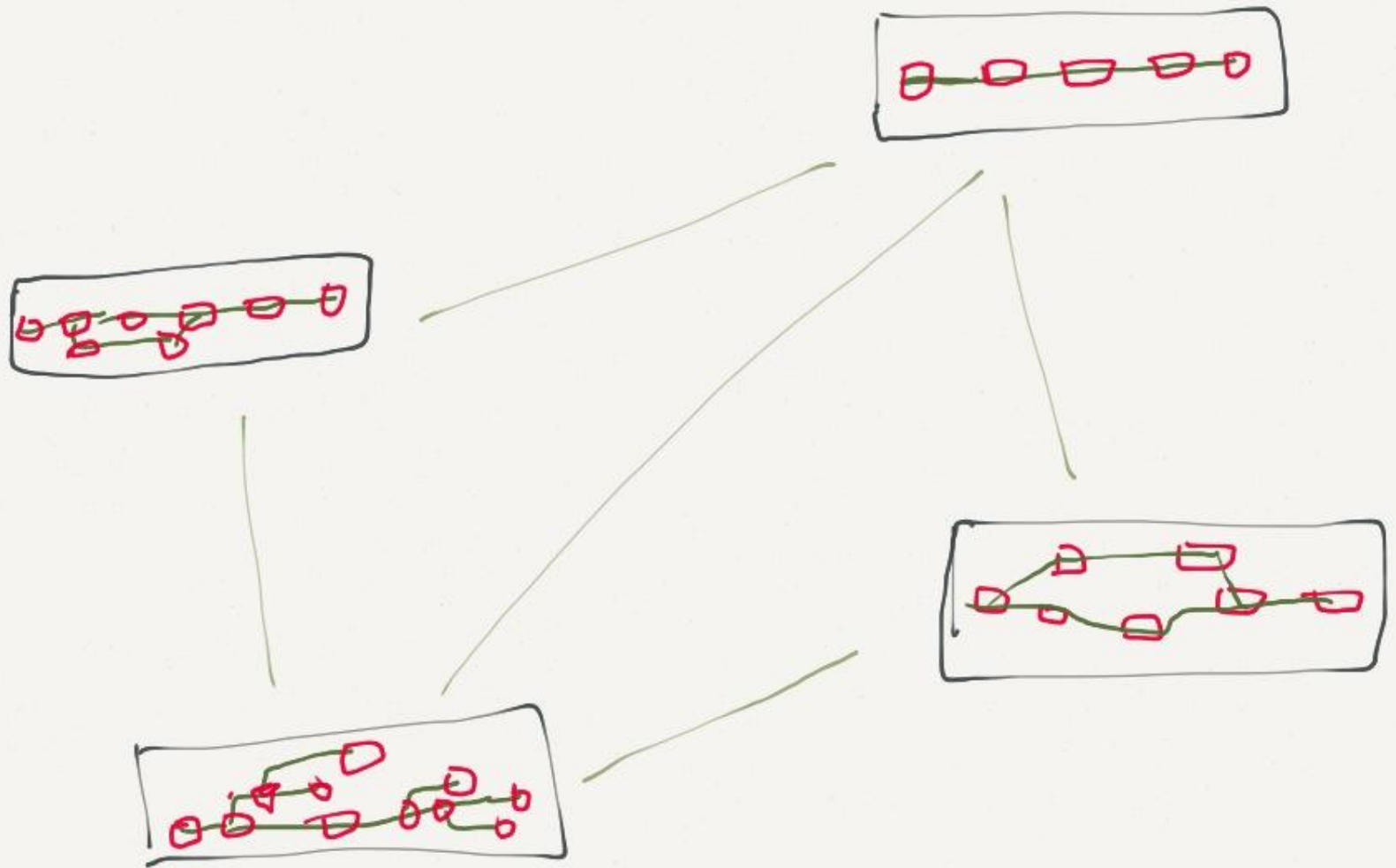`Commit` history provides a `log` of *punctuated meaningful work*
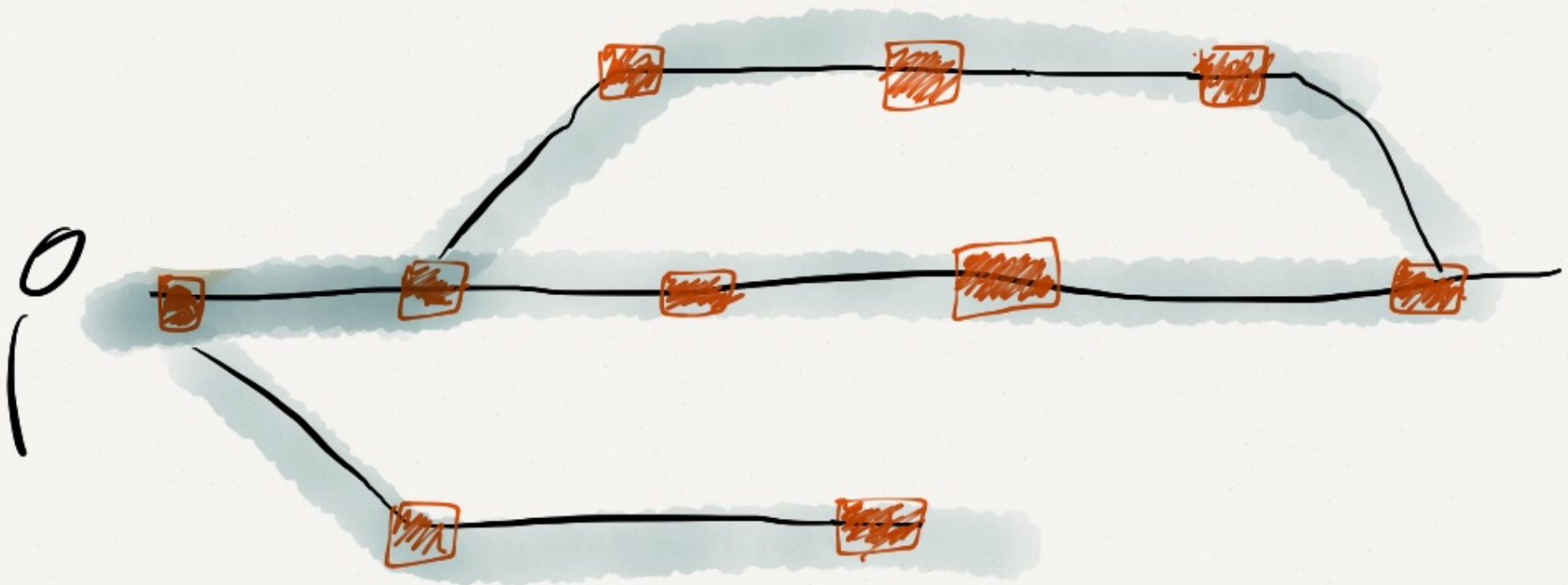
Branching & **Merging** these units enables parallel workflows.

Cloning or **Forking** these branches enables asynchronous remote collaboration.

Merging lets you incorporate branches or **Pull** remote work.

Git is a *protocol* and a *scaffolding* for sharing and scaling
Repositories.

While Git is technically indifferent to the *stuff* in the repository. The design favors structured text (code, html, xml, etc)

```
66  66            <h1>nPLSA Topic Model Vizualizer</h1>
67      -            <a class="vizlink" href="map.html">
68      -              <img class="img-rounded" src="img/topicmap.png">
69      -              <div class="caption">
70      -                <h3>Topic Map<h3>
71      -              </div>
72      -            </a>
73  67            <a class="vizlink" href="size.html">
74  68              <img class="img-rounded" src="img/topicsize.png">
75  69              <div class="caption">
76  70                <h3>Topic Size<h3>
77  71              </div>
78  72            </a>
    73  +            <a class="vizlink" href="map.html">
    74  +              <img class="img-rounded" src="img/topicmap.png">
    75  +              <div class="caption">
    76  +                <h3>Topic Map<h3>
    77  +              </div>
    78  +            </a>
79  79
80  80          </div> <!-- span10 -->
81      -        <div clas="span2">
    81  +        <div class="span5">
82  82            <h2>What is this?</h2>
83      -          <p> This site is a tool for analyzing the nPLSA topic modeling process.</p>
    83  +
    84  +<p>This site is a tool for analyzing the nPLSA topic modeling process. nPLSA or nonparametric Probabilistic
    Jian Tang, Ming Zhang, and Qiaozhu Mei. The nPLSA process is unique from other more popular topic modeling t
    Allocation (LDA) because it does not require you to specify the number of topics <em>a priori.</em> nPLSA in
    topics from documents in the corpus. For more information about nPLSA please read the paper on Arxive.org [l
    85  +
    86  +<p>The purpose of this tool is to being providing insight into the iterative process by visualizing the eme
    process. Currently the tool supports two kinds of interactive visualizations:</p>
```

Diff allows you to *see* the differences between **Commits**.

Git is great with **data,** not as great (but still good) with **prose.**
Need prose-centric vs. code-centric `diff` and `merge` ([Juxta](#) perhaps?)

Adding the June 2011 batch release

mcburton authored 6 months ago   `latest commit 14937ffa32`

..

📄 K000039.000.xml	6 months ago	Adding the February 2008 batch release [mcburton]
📄 K000122.000.xml	6 months ago	Adding the September 2008 batch release [mcburton]
📄 K000152.000.xml	6 months ago	Adding the September 2006 batch release [mcburton]
📄 K000180.000.xml	6 months ago	Adding the September 2009 batch release [mcburton]
📄 K000266.000.xml	6 months ago	Adding the February 2008 batch release [mcburton]
📄 K000268.000.xml	6 months ago	Adding the September 2006 batch release [mcburton]
📄 K000335.000.xml	6 months ago	Adding the June 2011 batch release [mcburton]
📄 K000343.000.xml	6 months ago	Adding the June 2011 batch release [mcburton]
📄 K000379.000.xml	6 months ago	Adding the September 2006 batch release [mcburton]
📄 K000406.000.xml	6 months ago	Adding the June 2011 batch release [mcburton]
📄 K000415.000.xml	6 months ago	Adding the September 2008 batch release [mcburton]
📄 K000454.000.xml	6 months ago	Adding the September 2009 batch release [mcburton]

Repository structure is complicated. Directories of XML files is not ideal.

Git and GitHub can be used as a platform for hosting, sharing, and tracking Digital Humanities data repositories.

mcburton 5 months ago Adding the February 2008 batch release

1 contributor

executable file | 3697 lines (3695 sloc) | 150.784 kb

Edit | Raw | Blame | History

```
1   <?xml version="1.0" encoding="utf-8"?>
2   <!DOCTYPE ETS SYSTEM "http://www.lib.umich.edu/tcp/docs/code/eebo2prf.xml.dtd">
3   <ETS>
4   <TEMPHEAD>
5   <REVDESCR>
6   <CHANGE>
7   <DATE>2007-03-19</DATE>
8   <RESPSTMT>
9   <NAME>Simon Charles</NAME>
10  <RESP>MURP</RESP>
11  </RESPSTMT>
12  <ITEM>Proofed and reviewed</ITEM>
13  </CHANGE>
14  </REVDESCR>
15  </TEMPHEAD>
16  <EEBO>
17  <IDG S="marc" R="OX" ID="K000039.000">
18  <STC T="C">N000071</STC>
19  <BIBNO T="docno">CW3306500660</BIBNO>
20  <VID SET="ECFA">1105700300</VID>
21  </IDG>
22  <TEXT LANG="eng">
23  <FRONT>
24  <DIV1 TYPE="half title">
25  <PB REF="1"/>
26  <P>A COMEDY, CALL'D, THE Basset-Table.</P>
```

Github as a host for DH data. ECCO-TCP is already there.

Github can be used as a central coordination point for *distributed, asynchronous collaboration.*
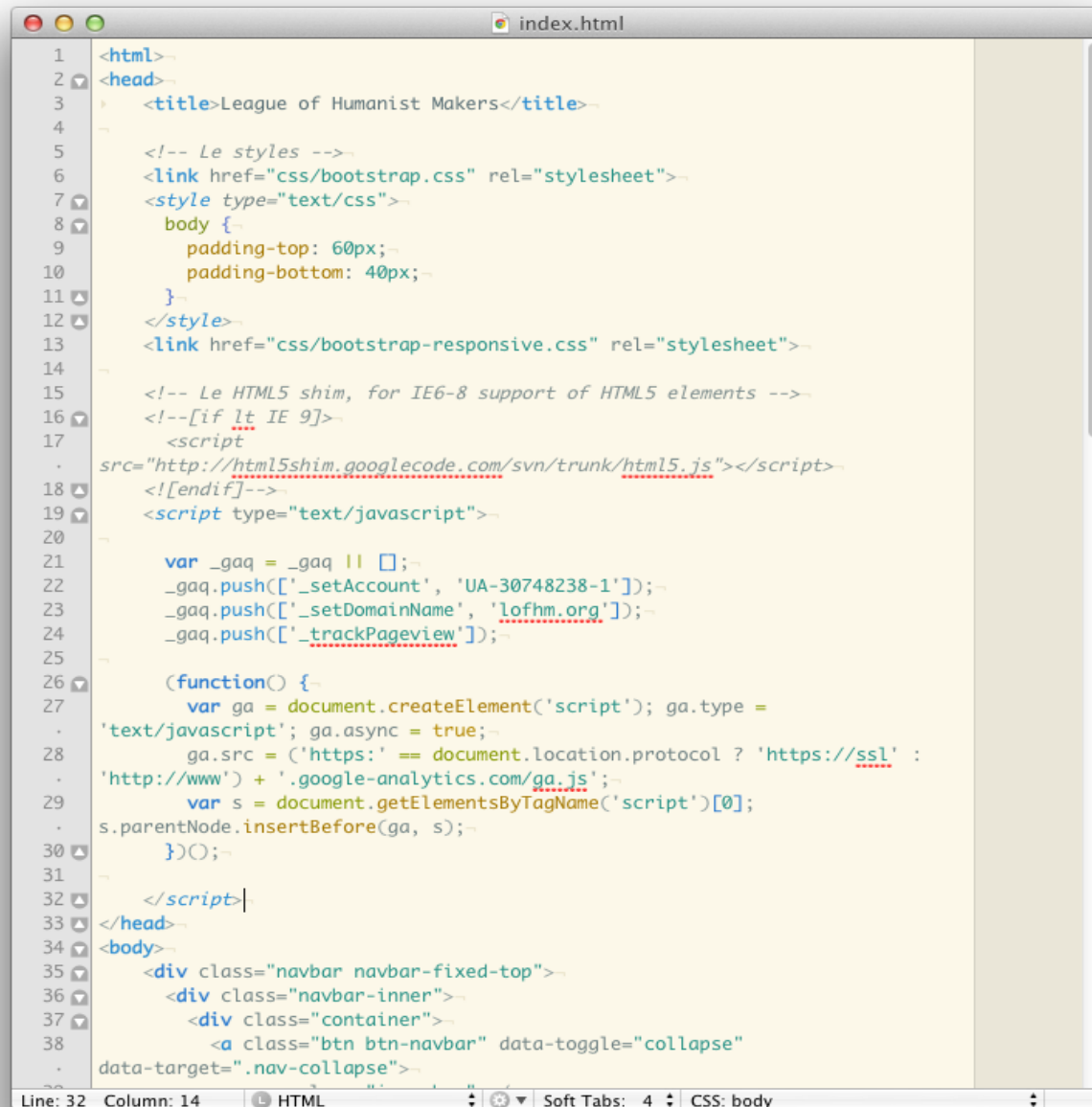
I can **Pull** a remote repository from Github to my local machine...

```
      1   <html>
      2   <head>
      3       <title>League of Humanist Makers</title>
      4
      5       <!-- Le styles -->
      6       <link href="css/bootstrap.css" rel="stylesheet">
      7       <style type="text/css">
      8         body {
      9           padding-top: 60px;
     10           padding-bottom: 40px;
     11         }
     12       </style>
     13       <link href="css/bootstrap-responsive.css" rel="stylesheet">
     14
     15       <!-- Le HTML5 shim, for IE6-8 support of HTML5 elements -->
     16       <!--[if lt IE 9]>
     17         <script
          src="http://html5shim.googlecode.com/svn/trunk/html5.js"></script>
     18       <![endif]-->
     19       <script type="text/javascript">
     20
     21         var _gaq = _gaq || [];
     22         _gaq.push(['_setAccount', 'UA-30748238-1']);
     23         _gaq.push(['_setDomainName', 'lofhm.org']);
     24         _gaq.push(['_trackPageview']);
     25
     26         (function() {
     27           var ga = document.createElement('script'); ga.type =
          'text/javascript'; ga.async = true;
     28           ga.src = ('https:' == document.location.protocol ? 'https://ssl' :
          'http://www') + '.google-analytics.com/ga.js';
     29           var s = document.getElementsByTagName('script')[0];
          s.parentNode.insertBefore(ga, s);
     30         })();
     31
     32       </script>
     33   </head>
     34   <body>
     35       <div class="navbar navbar-fixed-top">
     36         <div class="navbar-inner">
     37           <div class="container">
     38             <a class="btn btn-navbar" data-toggle="collapse"
          data-target=".nav-collapse">
```

Line: 32   Column: 14   HTML   Soft Tabs: 4   CSS: body

....edit files locally with my favorite editors and tools...

...**Commit** the changes to my local repository...
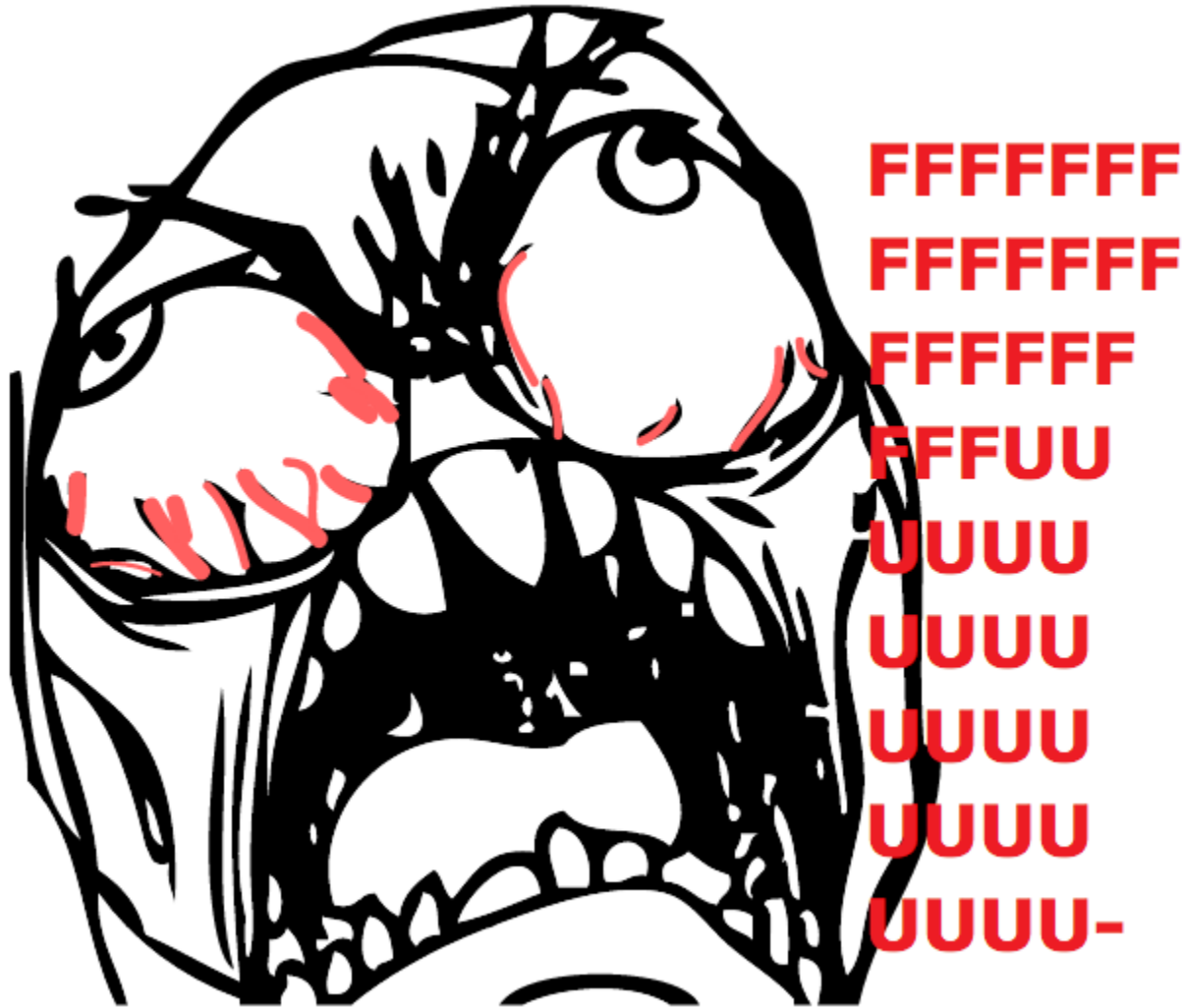
...and **Push** my updated repository to *GitHub*.

FFFFFF
FFFFFFF
FFFFFF
FFFUU
UUUU
UUUU
UUUU
UUUU
UUUU-

This is a *very* high level, conceptual overview.
Using Git is complicated and requires *lots* of patience.

Thanks! If you have questions please contact me on twitter (@mcburton) or email (matt@mcburton.net).