



# Be More {Productive, Collaborative} with {Git, GitHub}

---

Robert McGibbon, Kyle Beauchamp  
SimTk Meeting  
October 16, 2012

# git :: overview

---

# git :: overview

---

- (distributed) version control system
  - cvs, svn, hg, bzt

# git :: overview

---

- (distributed) version control system
  - cvs, svn, hg, bzt
- support for distributed, non-linear development

# git :: overview

---

- (distributed) version control system
  - cvs, svn, hg, bzt
- support for distributed, non-linear development
- originally developed for managing the linux kernel by linus torvalds

# git :: overview

---

- (distributed) version control system
  - cvs, svn, hg, bzz
- support for distributed, non-linear development
- originally developed for managing the linux kernel by linus torvalds
- extremely flexible, can be complicated

# git :: simple workflow

---

```
# make a local copy of a repository  
$ git clone <url>  
$ vi new_module.py
```

```
# stage your work  
$ git add new_module.py
```

```
# commit early and often -- local  
$ git commit -m "Sweet new feature!"
```

```
# upload your changes to the remote  
$ git push
```

# github :: overview

---

Fork me on GitHub



# github :: overview

---

- collaborative/“social” code hosting web app
  - used by facebook, twitter, mozilla, ...
  - hosts numpy, scipy, ipython, matplotlib, ...

# github :: overview

---

- collaborative/“social” code hosting web app
  - used by facebook, twitter, mozilla, ...
  - hosts numpy, scipy, ipython, matplotlib, ...
- center of the open source “do-ocracy”
  - everything is forkable

# github :: overview

---

- collaborative/“social” code hosting web app
  - used by facebook, twitter, mozilla, ...
  - hosts numpy, scipy, ipython, matplotlib, ...
- center of the open source “do-ocracy”
  - everything is forkable
- currently trendy; \$100M VC money

# github :: collaborative

---

Fork me on GitHub

# github :: collaborative

---

- pull requests
  - accept contributions from the community without giving up control.
  - manage concurrent development across features / branches

# github :: collaborative

---

- pull requests
  - accept contributions from the community without giving up control.
  - manage concurrent development across features / branches
- issue tracker
  - joint developer to-do list, bug tracking system

# github :: collaborative

---

- pull requests
  - accept contributions from the community without giving up control.
  - manage concurrent development across features / branches
- issue tracker
  - joint developer to-do list, bug tracking system
- interface focuses on the code and the users

# github :: pull requests

The screenshot shows a web browser window with multiple tabs open. The active tab is a GitHub pull request for the repository `SimTk/msmbuilder`. The URL in the address bar is `https://github.com/SimTk/msmbuilder/pull/39`. The GitHub logo and navigation links are visible at the top. The repository name `SimTk/msmbuilder` is displayed, along with buttons for `Pull Request`, `Unwatch`, `Star` (4), and `Fork` (1). Below the repository name, there are tabs for `Code`, `Network`, `Pull Requests` (1), `Issues` (13), `Wiki`, `Graphs`, and `Admin`. The `Pull Requests` tab is selected. A summary bar indicates that the pull request is `Closed`, created by `rmcgibbo`, and aims to merge 37 commits into the `master` branch from the `msmio` branch. It shows 3,898 additions and 1,615 deletions, and is labeled `#39`. Below this, there are tabs for `Discussion`, `Commits` (37), and `Files Changed` (56). The main content area shows the pull request details. It starts with a message from `rmcgibbo` stating they opened the pull request a month ago. The title of the pull request is `Serializer replacement / rewrite of Project`. Below the title, it says "No one is assigned" and "No milestone". The description of the pull request is as follows: 

@kyleabeauchamp: thoughts?

1. Replaces the `Serializer.py` file with `msmio.py`. `msmio.py` contains two functions, `saveh` and `loadh` that save and load numpy arrays to hdf5 files. They have roughly the same interface as `numpy.savez` and `numpy.load`.
2. Replaces `Project` with a version that loads and saves from flat text.

Below the description, it shows 5 participants. At the bottom, there is a commit history section showing that `rmcgibbo` added some commits a month ago. The commit message is `(np.savez/np.load)-like Serializer replacement` with the commit hash `4f2f1f0`.



# github :: pull requests

 schwancr started a discussion in the diff 4 days ago

TestFiles/TestWrappers.py [View full changes](#)

```
... @@ -181,16 +179,16 @@ def test_f_CalculateImpliedTimescales(self):
188 -     # make a predictable stream of random numbers by seeding the RNG with 42
189 -     random_source = np.random.RandomState(42)
190 -     randomconfs = GetRandomConfs.run(P1, Assignments, NumRandomConformations, random_source)
191 -
192 -     reference = Trajectory.load_trajectory_file(os.path.join(ReferenceDir, "2RandomConfs.lh5"))
193 -     self.assert_trajectories_equal(reference, randomconfs)
182 + # def test_g_GetRandomConfs(self):
183 + #     P1 = Project.load_from(ProjectFn)
184 + #     Assignments = io.loadh("Data/Assignments.Fixed.h5", 'arr_0')
185 + #
186 + #     # make a predictable stream of random numbers by seeding the RNG with 42
187 + #     random_source = np.random.RandomState(42)
188 + #     randomconfs = GetRandomConfs.run(P1, Assignments, NumRandomConformations, random_source)
189 + #
190 + #     reference = Trajectory.load_trajectory_file(os.path.join(ReferenceDir, "2RandomConfs.lh5"))
191 + #     self.assert_trajectories_equal(reference, randomconfs)
```

2

 schwancr [repo collab](#) 4 days ago  

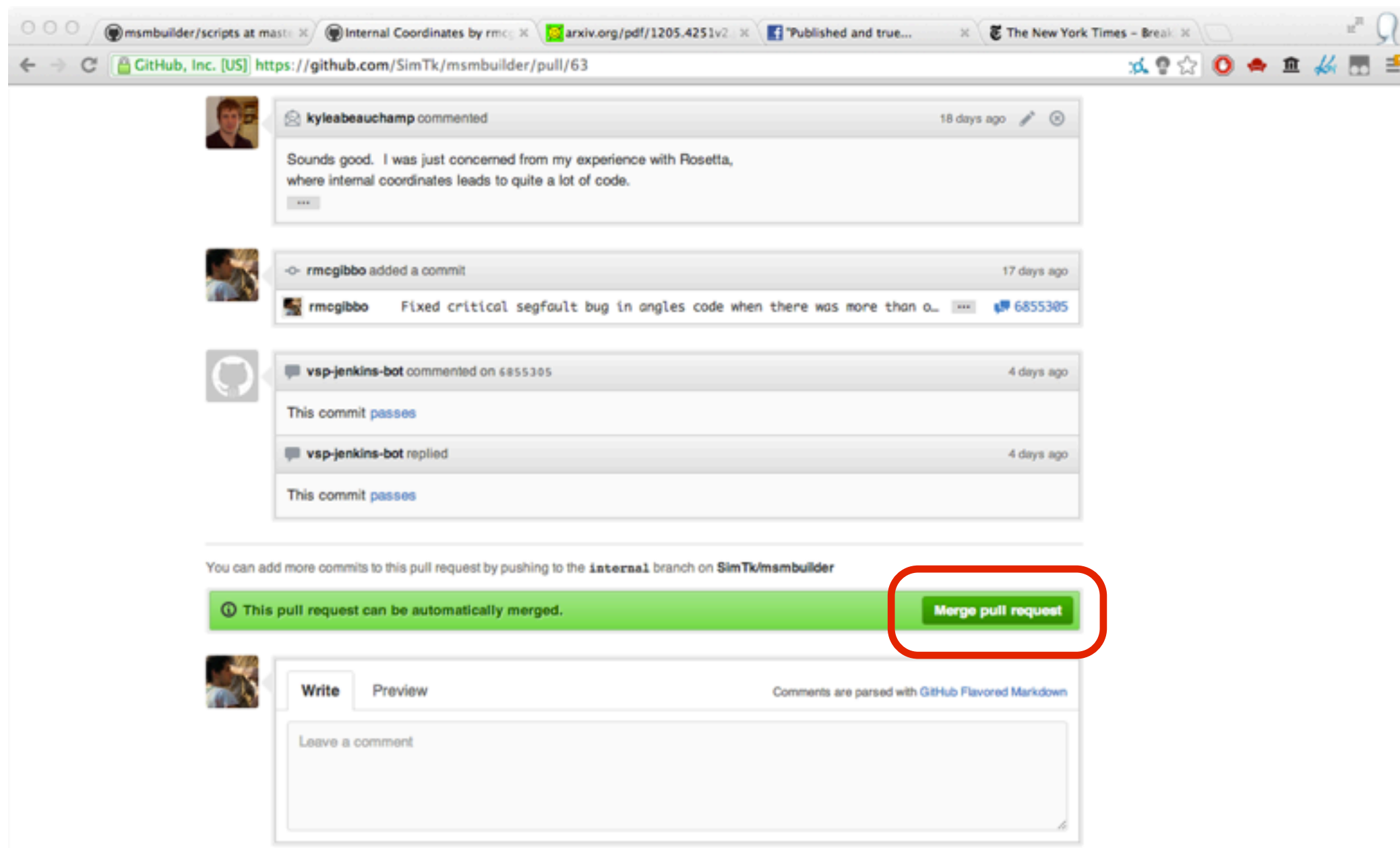
Why are these commented out?

 rmcgibbo [repo collab](#) 4 days ago  

Because I was lazy, and when I deleted GetRandomConfs.py I didn't readjust them.

[Add a line note](#)

# github :: pull requests



# github :: issues

The screenshot displays the GitHub issues interface for a repository. At the top, there are tabs for '13 open issues' and '59 closed issues'. To the right, there are filters for 'Submitted', 'Updated', and 'Comments'. Below these are filters for 'Close', 'Label', 'Assignee', and 'Milestone'. The list of issues is as follows:

Issue Number	Title	Labels	Author	Time Ago	Comments
#71	Avoid copy/paste code for logging.	style	rmcgibbo	2 days ago	5 comments
#67	PDB Reader should preserve ResidueID		kyleabeauchamp	3 days ago	
#60	Implement history functionality in ProjectBuilder/ConvertDataToHDF.py		tjlane	3 days ago	
#41	Proposed changes to Trajectory	question	kyleabeauchamp	3 days ago	28 comments
#52	Make docstrings render nicely		rmcgibbo	3 days ago	2 comments
#16	MLE getting stuck for hours	question, possible bug	rmcgibbo	3 days ago	8 comments
#58	Trajectory Should Contain Timestep Metadata.	enhancement, question	tjlane	3 days ago	7 comments
#70	Write test for SaveStructures.py		rmcgibbo	3 days ago	
#63	Internal Coordinates		rmcgibbo	11 days ago	6 comments

Additional details for the last issue (#63): Code Attached

# github :: issues



rmcgibbo commented

25 days ago

@tjlane: Do you want to start hacking the fahproject stuff on this?

I think the `validators` mechanism here will generalize pretty well for the FaH stuff.

basically, you do

```
pb = ProjectBuilder(...)
pb.add_validator(validators.RMSDExplosionValidator(structure_or_filename='native.pdb', max_rmsd=0.1))

project = pb.get_project()
project.save('ProjectInfo.yaml')
```



tjlane commented

25 days ago

Sweet! Nice work. Especially like the Validators -- I wouldn't worry about adding more, just documenting how they work so people can add them if necessary.

I'll get cracking on this in a bit.

# github :: beautiful

---

Fork me on GitHub

# github :: beautiful

---

- the little things make it useful
  - low friction, lightweight “project management”
  - code highlighting
  - diff discussion
  - reply to comments via email
  - web hooks

# github :: extensible

---

Fork me on GitHub

# github :: extensible

---

- continuous integration
  - run the build/test suite after every commit across a matrix of configurations
  - TravisCI, CircleCI



# github :: extensible

---

- continuous integration
  - run the build/test suite after every commit across a matrix of configurations
  - TravisCI, CircleCI
- documentation generation and hosting
  - build the documentation, host it in the cloud
  - ReadTheDocs.org

# github :: readthedocs.org

Fork me on GitHub

The screenshot shows a web browser window with the URL `msmbuilder.readthedocs.org/en/latest/index.html` highlighted in the address bar. The page title is "MSMBuilder 2.6.dev documentation". The main content area is titled "Welcome to MSMBuilder's documentation!" and lists the following contents:

- MSMBuilder Introduction
- Distance Metrics: `msmbuilder.metrics`
  - Performance Notes
  - Kinetic Distance Metrics
  - Concrete Metrics
  - Abstract Classes
  - Utility Methods
- Clustering: `msmbuilder.clustering`
  - Abstract Classes
  - Flat Clustering Classes
  - Hierarchical Clustering
  - Clustering Functions
  - Utility Functions
- assigning: `msmbuilder.assigning`
  - Functions
- MSM Construction: `msmbuilder.MSMLib`
  - Notes
  - Mapping
  - Trimming
  - Model Building
- MSM Analysis: `msmbuilder.msm_analysis`
  - Notes
  - Model Queries

The left sidebar contains a logo with the letters "MSM" and a crane, a "Project Versions" section showing "latest", an "RTD Search" box with a "Go" button, a "Table Of Contents" section with links to "Welcome to MSMBuilder's doc", "Indices and tables", a "Next topic" section with a link to "MSMBuilder Introduction", and a "This Page" section with a link to "Show Source".

# github :: travis-ci.org

The screenshot shows the Travis CI web interface for the `matplotlib/matplotlib` repository. The interface is divided into several sections:

- Left Sidebar:** A list of recent builds. Each entry includes a status icon (yellow for pending, green for successful), the repository name, build number, duration, and completion status.
 

Repository	Build #	Status	Duration	Finished
robotex82/ecm_cms	#13	Pending	1 min 3 sec	-
lyrix/Silex-Kitchen-Edition	#73	Pending	1 min 13 sec	-
yahoo/mojito	#488	Pending	1 min 56 sec	-
tulsawebdevs/okletsgo	#29	Success	47 sec	about a minute ago
rolandwalker/font-utils	#29	Success	3 min 23 sec	about a minute ago
Limesco/CServ	#14	Success	2 min 10 sec	about a minute ago
brettweavnet/cucumber-loggly	#3	Success	45 sec	3 minutes ago
- Center Panel:** Detailed view of the current build (442). It includes tabs for Current, Build History, Pull Requests, and Branch Summary. The build status is pending (yellow circle). The commit is `8df51a7 (v1.2.x)`. The build duration is 11 min 19 sec. The message is "Remove unused autolayout\_validator() from rcsetup".
 

Job	Duration	Finished	Python
442.1	6 min 29 sec	5 minutes ago	2.6
442.2	8 min 50 sec	2 minutes ago	2.7
442.3	11 min 19 sec	-	3.1
442.4	10 min 21 sec	-	3.2
- Right Sidebar:** Contains a "Sponsors" section with logos for mongoHQ and kanbanery, a "Workers" section listing various worker nodes, and a "Queue: Common" section showing a list of builds waiting to be executed.

github :: simtk.org synergy

---

Fork me on GitHub

# github :: simtk.org synergy

---

Fork me on GitHub

- source code management is not our speciality; we can leverage external tools

# github :: simtk.org synergy

---

- source code management is not our speciality; we can leverage external tools
- simtk.org and github have distinct and complementary core competencies
  - simtk.org as project “front end” to centralize and advertise
  - github as a “back end” for source control and developer tools







# git :: criticisms

---

I didn't really expect anyone to use [Git] because it's so hard to use, but that turns out to be its big appeal. No technology can ever be too arcane or complicated for the black t-shirt crowd. - Fake Interview with Linus Torvalds