Submodule Pain
- or How I Learned to
Stop Worrying and
Love the Subtree

Jackson Oursland



#### git Submodules

- git submodules are a means of pointing to an external repository.
- Submodules have a few components:
  - .gitmodules entry with the remote repo URL and path to check out
  - .git/modules/ directory which stores the .git of the module
  - reference in the index and objects which points to the SHA of the commit to be checked out
- Note that the files themselves are NOT tracked, only the commit reference which is typically a SHA.

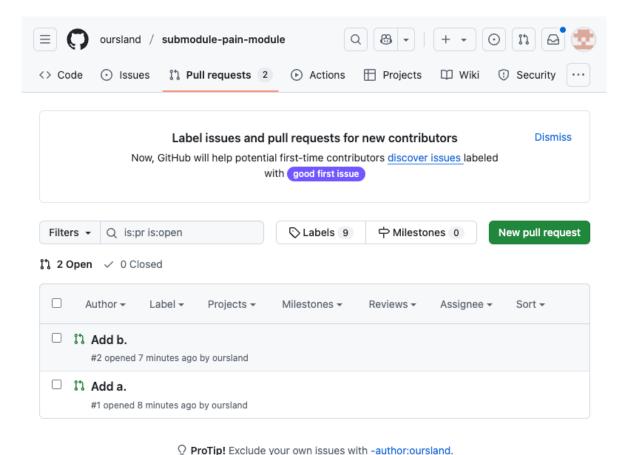
#### Reference Repositories

- oursland/submodule-pain
  - Top-level repo
- oursland/submodule-pain-module
  - Submodule

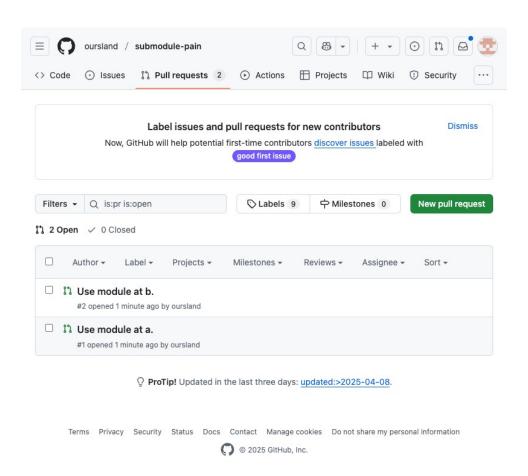
#### PRs Take One: Branches

- oursland/submodule-pain-module
  - PR#1: Branch 'a' adds the file 'a'
  - PR#2: Branch 'b' adds the file 'b'
  - There are no possible conflicts
- oursland/submodule-pain
  - Branch 'master'
    - 81b3d609d3158822a6a37966436f5c64e1eb4ac6 module (heads/master)
  - PR#1: Branch 'a' updates the submodule at branch 'a'
    - c7c5d8c0d28bade1ba45706bf49d85df57771947 module (heads/a)
  - PR#2: Branch 'b' updates the submodule at branch 'b'
    - 75b743cf403b92c8f631d419490f04eef9832d1b module (heads/b)

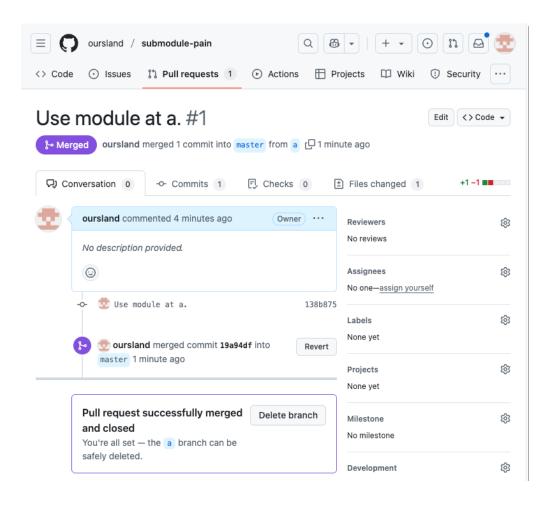
# Submodule repo PRs



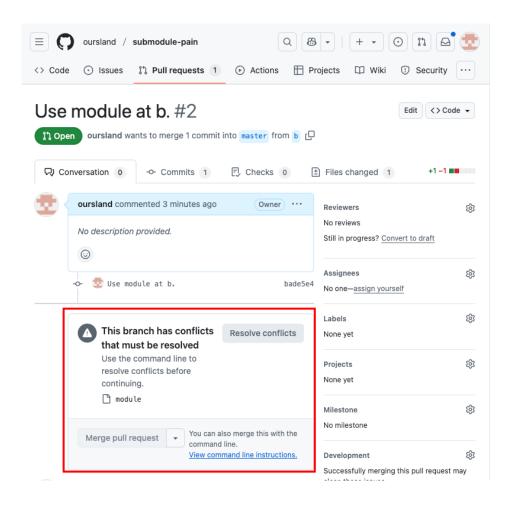
# Top-level repo PRs



# Merge 'a' into Top-Level Repo



# Now the 'b' PR cannot merge due to "conflicts"



#### Recap

- No actual conflicts are present.
- One PR added the file 'a' and the other added the file 'b', but because they are submodules git has no knowledge of the files or their compatibility.
- Because each PR uses a different SHA, then the merge cannot take place because the conflict is in the SHA values.
  - PR#1: 81b3d609d3158822a6a37966436f5c64e1eb4ac6 -> c7c5d8c0d28bade1ba45706bf49d85df57771947
  - PR#2: 81b3d609d3158822a6a37966436f5c64e1eb4ac6 -> 75b743cf403b92c8f631d419490f04eef9832d1b

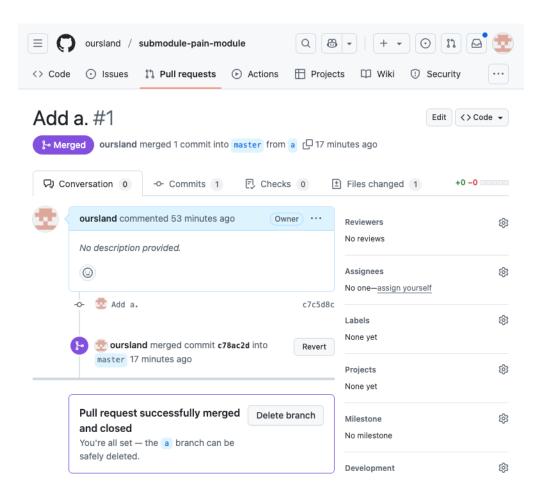
#### PRs Take Two: Default Branch Only

- Okay, normally we do not want people to have submodules at branches which may not have been reviewed, but rather to only accept changes from:
  - tags
  - commits on default branch
  - commits on a release tracking branch
- Let's fix that with two new PRs
  - Based upon the current `master`, so 'module' now points to 'a'
  - PR#3: adds the file 'a', but after being merged into 'master'
  - PR#4: adds the file 'b', but after being merged into 'master'

#### PRs Take Two: Default Branch Only

- oursland/submodule-pain-module
  - Adds the file 'a' after merging PR#1 into submodule
  - Adds the file 'b' after merging PR#2 (both 'a' and 'b' are present)
- oursland/submodule-pain
  - PR#3: Branch 'a2' updates the submodule at branch 'master' after PR#1
    - c78ac2db2eb6ca7df9877249135affe0411cf853 module (heads/master)
  - PR#4: Branch 'b2' updates the submodule at branch 'master' after PR#2
    - 5f015926eae1ba7100d6fb34c46b8e18b10ed5ec module (heads/master)

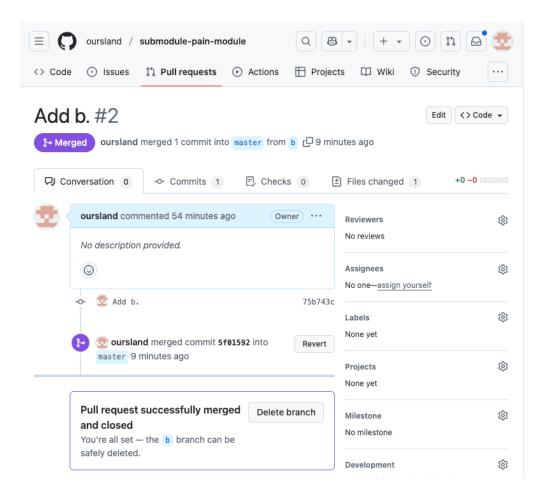
## Merge PR#1 into Submodule



## Create PR#3 in Top-Level Project

- Submodule now has 'a' in the 'master' branch after code-review
- Submit PR#3 with the updated submodule

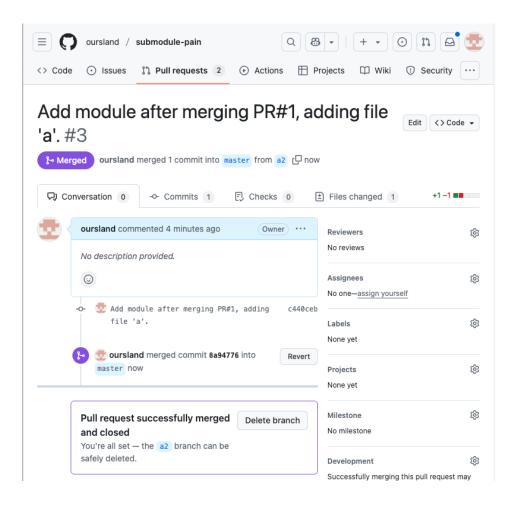
## Merge PR#2 into Submodule



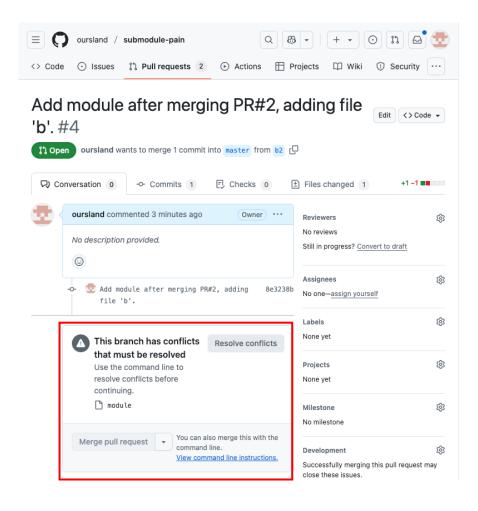
## Create PR#4 in Top-Level Project

- Submodule now has 'b' in the 'master' branch after code-review
- Submodule now has both 'a' and 'b'
- Submit PR#4 with the updated submodule

# Merge PR#3 into Top-Level



# Cannot merge PR#4 due to "conflicts"



# **Again Conflicts**

- As before, git submodules do not track actual code changes and cannot perform merges
- The submodule's reference has changed in multiple PRs. After the first is merged, every following PR that touches that module will require manual resolution
- Increasing the number of submodules increases the likelihood of these conflicts

#### How this affects FreeCAD

- Proliferation of submodules increases the likelihood that a PR will cause a cascade of merge conflicts requiring manual resolution
- Even simple changes like mosfet80's frequent code cleanup PRs will invalidate every PR that takes place later in the chain
- Merge Request Mondays could be brought to a halt
- As davesrocketshop can attest to, weekly manual conflict resolution is no fun and does not push the project forward

#### How to Solve This?

- Let git do what it is meant to do
  - Track source code in a line-oriented way, not SHAs
  - Perform automatic merges when changes do not conflict
- Check in the source
  - Vendoring a release is an option, but it makes tracking local changes challenging and submitting patches upstream harder
  - Branches for vendored releases can be merged, which make managing the local repo easier, but still does not make submitting patches upstream any easier

# How to Solve This? (cont.)

- Check in the source (cont.)
  - Vendoring a remote repo is another option
    - Checkout a clone into the host repo (you can have multiple roots in a git repo!)
    - Use git read-tree to check-out the source at a given reference to a subdirectory
    - Use git filter-branch to reconstruct patches for submission upstream
    - git read-tree and git filter-branch are both arcane magick that should not be used casually. git filter-branch's man page even comes with a massive warning.
  - Use git subtree

## git Subtree

- git subtree add --prefix=path/to/subdirectory/ repo-url branch-name
  - Will check out the repo at the desired branch and place it into the subdirectory.
  - This is done with a merge commit, preserving the history of the files.
  - Adding --squash will squash all commits to a single commit that is merged.
  - The merge commit's message contains references to path of subtree, current repo when the subtree is added, and the remote repo when it was merged

```
Add 'src/Mod/Assembly/OndselSolver/' from commit
'09d6175a2ba69e7016fcecc4f384946a2f84f92d'

git-subtree-dir: src/Mod/Assembly/OndselSolver
git-subtree-mainline: 5eb5af7068ac9e610ddbb3c62f6a5a54c7c3f818
git-subtree-split: 09d6175a2ba69e7016fcecc4f384946a2f84f92d
```

## git Subtree Pulling in Upstream Changes

- git subtree pull --prefix=path/to/subtree remote-repository branch-name
  - Will pull in changes from the upstream repo
  - The references in the prior merge commit are used to select which changes to bring
- git subtree push --prefix=path/to/subtree remote-repository branch-name
  - Will push changes to the upstream repository
  - References in the prior merge commit are used to identify commits that are to be pushed upstream

#### Removing a git Subtree

- Just run git rm -r path/to/subtree
  - The subtree is just normal code. The metadata is in the merge commit message.
- Contrast that to the git submodule deletion

```
git submodule deinit -f -- path/to/submodule
rm -rf path/to/submodule
git rm -f path/to/submodule
```

- Then commit
- Note: switching branches which still have the submodule can cause problems