

eAST usage of git-lfs

Git-lfs








- <https://git-lfs.github.com/>
- Store large files separately
 - Still versioned
 - Does not clog up your general storage and bandwidth
 - Smaller download until needed
- 1 GB included for free, EIC has 100 GB (for \$10/month, I think?)
 - Only total use is monitored; GUI doesn't say file X is using Y GB
 - Not sure what "100 GB bandwidth" means in practice

Tests

- Two identical files, one with lfs tracking, the other without
- No advantage is alive in the active branch(es)
 - except for quota if using external storage
 - in fact, **disadvantage**: not zipped (bad for e. g. gdml, pdf, csv)
- `git rm` on the active branch:
 - regular file's repo remains large (b/c of history)
 - git-lfs file is **only downloaded if switching to a commit** with the file
- Transparent to the user unless it fails
- Forks retain only the pointer to the original location
 - **Files deleted on the original repo stay lost in forks**

eAST Data → <https://github.com/eic/EicData>

- Transferred all data, with creation/extraction macros

	kkauder Added .mat and .Bmap files; updated RE... ✓ 0cbf467 1 hour ago 🕒 6 commits	
 ATHENA	Added ATHENA files	4 days ago
 Beampipe	Added .mat and .Bmap files; updated README	1 hour ago
 ECCE	Added .mat and .Bmap files; updated README	1 hour ago
 Solenoid	Added .mat and .Bmap files; updated README	1 hour ago
 .gitattributes	Added .mat and .Bmap files; updated README	1 hour ago
 README.md	Added .mat and .Bmap files; updated README	1 hour ago

Submodule

- When cloning eAST, these subdirectories are empty.
- To populate:
 - `git submodule init; git submodule update` or
 - `git clone --recurse-submodules [...]`
- .gitmodule should use the https url; one can use `git config submodule.EicData.url git@github.com:eic/EicData.git` to keep pushing with ssh
- We have the best use case, but...

If there are problems, we could try
subtrees, subrepos



Last Slide

- To Do:
 - Cleanup: small documentation stages, fix run.mac
 - Pages can incorporate submodule documentation in some sense, not in our case. Bit awkward to have to touch the main repo to document changes to the data one, but no big deal
- Questions:
 - Zip the data? Probably 5x size reduction
 - More submodules? Users now need to download 240 MB just for run.mac
 - Clean large-ish files from history?
I've done it before but would appreciate help