



# Cleaning up larsim Or 4 Weird Things you Probably Didn't Know About Git

Marc Mengel -- Scientific Data Management Group

# Overview

- Situation
- 4 Weird Things...
- How it happens
- Cleanup options

# Current Larsim Situation

- Multiple past add/commit/push of Gigabyte .root files
- Git clone of larsim → 11 G of disk
- Packing larsim → eats all memory on systems
- Breaks build machines
- Makes cdcvs even slower

## Four Weird Things...

- A “git clone” gives you a whole repository copy
  - Full history
  - Everything ever committed
- A “git push”
  - moves all those historical objects to the destination
- Revision Id's
  - are recursive checksums of whole directory
  - and prev revision
  - Changing 1 bit anywhere in history changes all later revision tags
- Git packes objects into compressed packfiles

# How it Happens

- You're working on a package that makes .root files
- Someone accidentally git-add's one or two and commits
- They git-remove it in a later revision (but its still in the history, in the repository...)
- Lather, rinse, repeat...
- Each iteration adds a Gig or so to the repository

# Cleanup Options

- Flag Day
  - Everyone commits/pushes/removes their clones
  - Scrub cdcvs copy
  - Checkout clean ones
- Incremental
  - Freeze incoming pushes
  - Scrub cdcvs copy
  - Folks clone, copy in their changes, remove their old clone
- Everybody cleans
  - Distribute incantation to remove .root files from all of history
  - Everybody runs it (eating all memory on their system for hours)

# Conclusion

- We need to pick a cleanup path
- Follow prevention measures
  - .gitignore
  - prereceive hook
- Policy
- Using git repositories in larsoft wiki