

CABLE testing infrastructure

Ned Haughton

March 16, 2016

The ONLY VALID MEASUREMENT
OF CODE QUALITY: WTFs/MINUTE

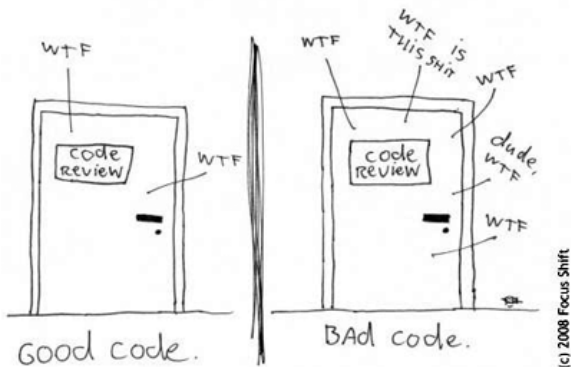


Figure 1: All code is bad. It's good to know how bad

Why automated testing?

- ▶ Ensure that code never gets worse (*regression testing*)
- ▶ Ensure code is doing what it's meant to do (*unit testing* - not always viable with science code)
- ▶ Maximise ease of maintenance (*code style tests*)
- ▶ Performance testing (science, efficiency)
- ▶ Automation saves time and drudge work, ensures that testing is consistent, and allows results to be safely ignored if they are “good”.

Types of testing

- ▶ Technical tests:
 - ▶ Compile/runs testing, with warnings enabled
 - ▶ Code style tests
 - ▶ Bitwise reproducibility on non-science changes
 - ▶ Unit tests where possible (e.g. on functions)
- ▶ Science tests:
 - ▶ various metrics, comparison with benchmarks
 - ▶ It's unlikely that a strict improvement in science code will improve all metrics, so
 - ▶ Subjective thresholds can be set (e.g. unacceptable, acceptable, good).

- ▶ Jenkins - open source continuous integration suite
 - ▶ Runs scripts on Raijin (e.g. build, run scripts, arbitrary analysis scripts)
 - ▶ Collects and neatly summarises results
 - ▶ Can track changes
 - ▶ Can run tests over multiple set-ups (different compilers, library versions, CPU architectures, if available)
 - ▶ <https://climate-cms.nci.org.au/jenkins/job/CABLE/>
- ▶ PALS (Protocol for the Analysis of Land Surface models)
 - ▶ Science testing
 - ▶ Should be possible to send and receive test data/analysis results between Jenkins and PALS.
 - ▶ Currently offline, under development

Ideal use-case

New user working on code, wants to submit a modification

- ▶ Ticket opened on Trac
- ▶ Branch created from trunk
- ▶ Changes pushed to branch
- ▶ User triggers tests on their branch
 - ▶ If test results return hard problems, user fixes their branch, and re-submits
 - ▶ If tests show improvement/no serious degradation, trac issue is marked as tests-pass (automated if above performance thresholds?)
- ▶ Issues marked as test-pass go to CABLE committee for human review
- ▶ Branch merged into trunk

Basically, this process just ensures that a whole lot of drudge work gets done that that would otherwise either have to be done manually, or would not get done at all.