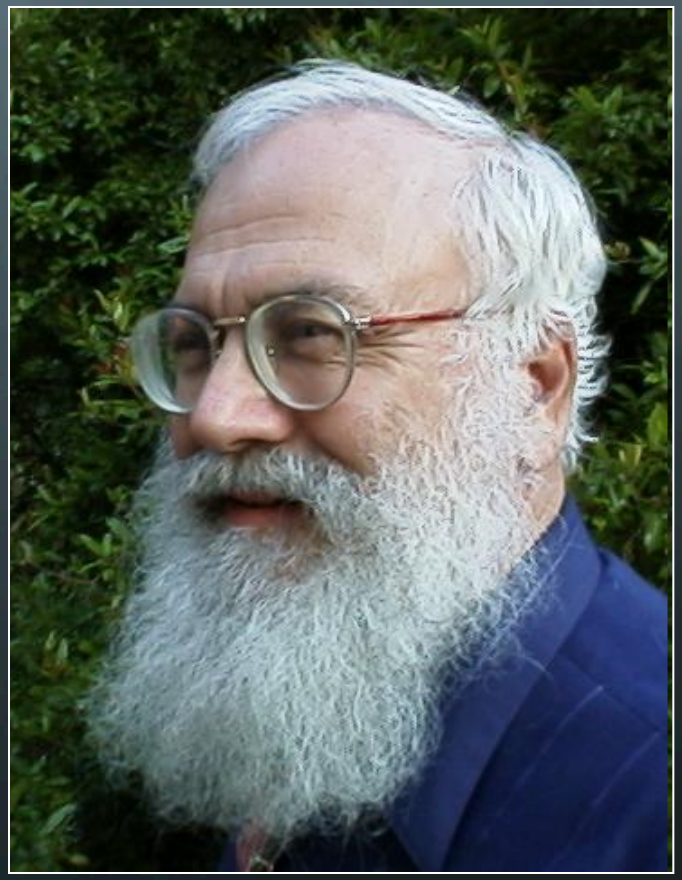
Sergey Fomel
Jackson School of Geosciences
The University of Texas at Austin

Workshop on Software Infrastructure for Reproducibility in Science
New York, May 30, 2013

# Claerbout's principle

"An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the **complete software development environment and the complete set of instructions which generated the figures**."

(Buckheit and Donoho, 1995)



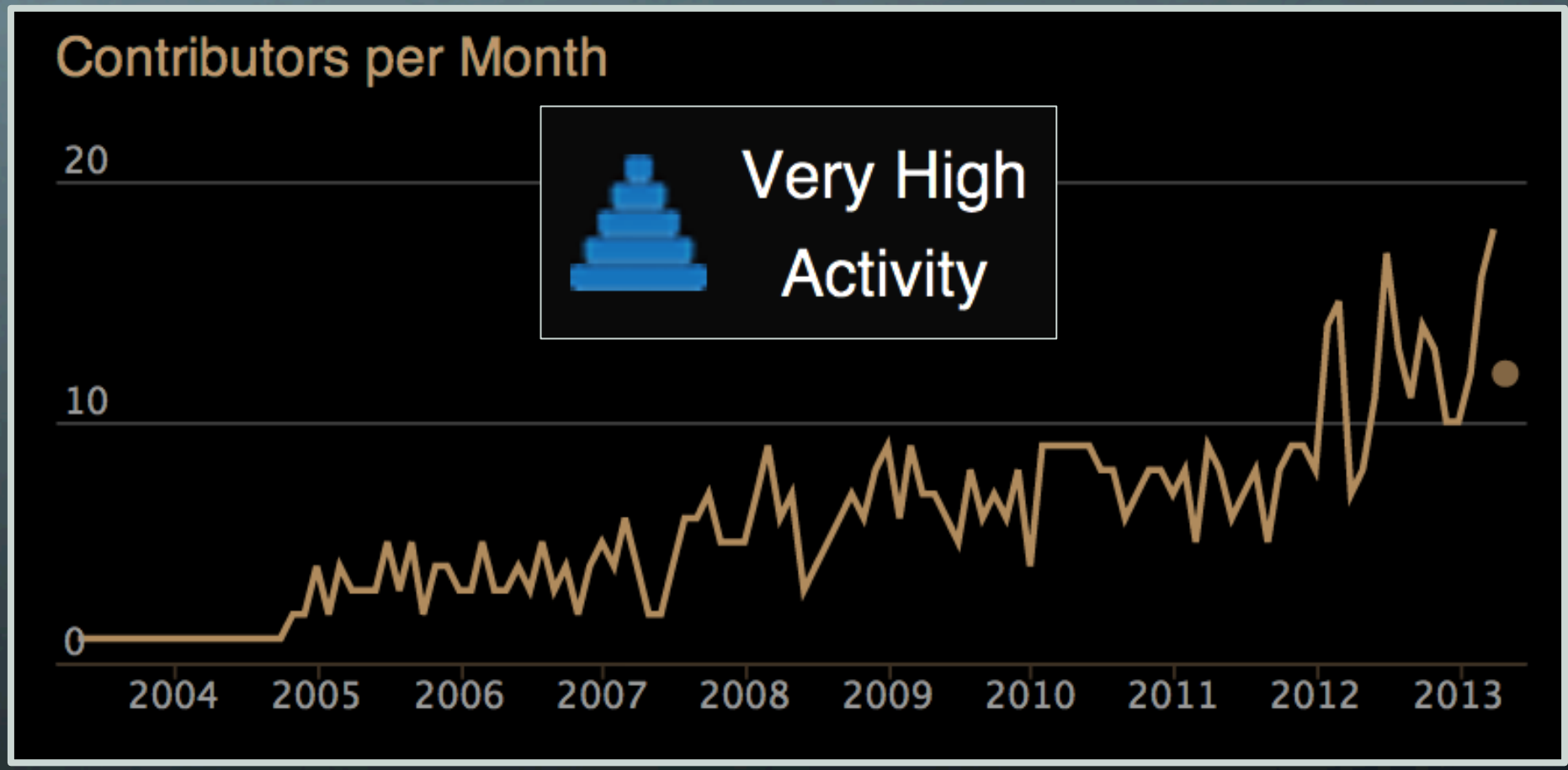
# Reproducible Research

"It is a big chore for one researcher to reproduce the analysis and computational results of another [...] I discovered that this problem has a simple technological solution: illustrations (figures) in a technical document are made by **programs and command scripts that along with required data should be linked to the document itself** [...] This is hardly any extra work for the author, but it makes the document much more valuable to readers who possess the document in electronic form because they are able to track down the computations that lead to the illustrations." (Claerbout, 1991)

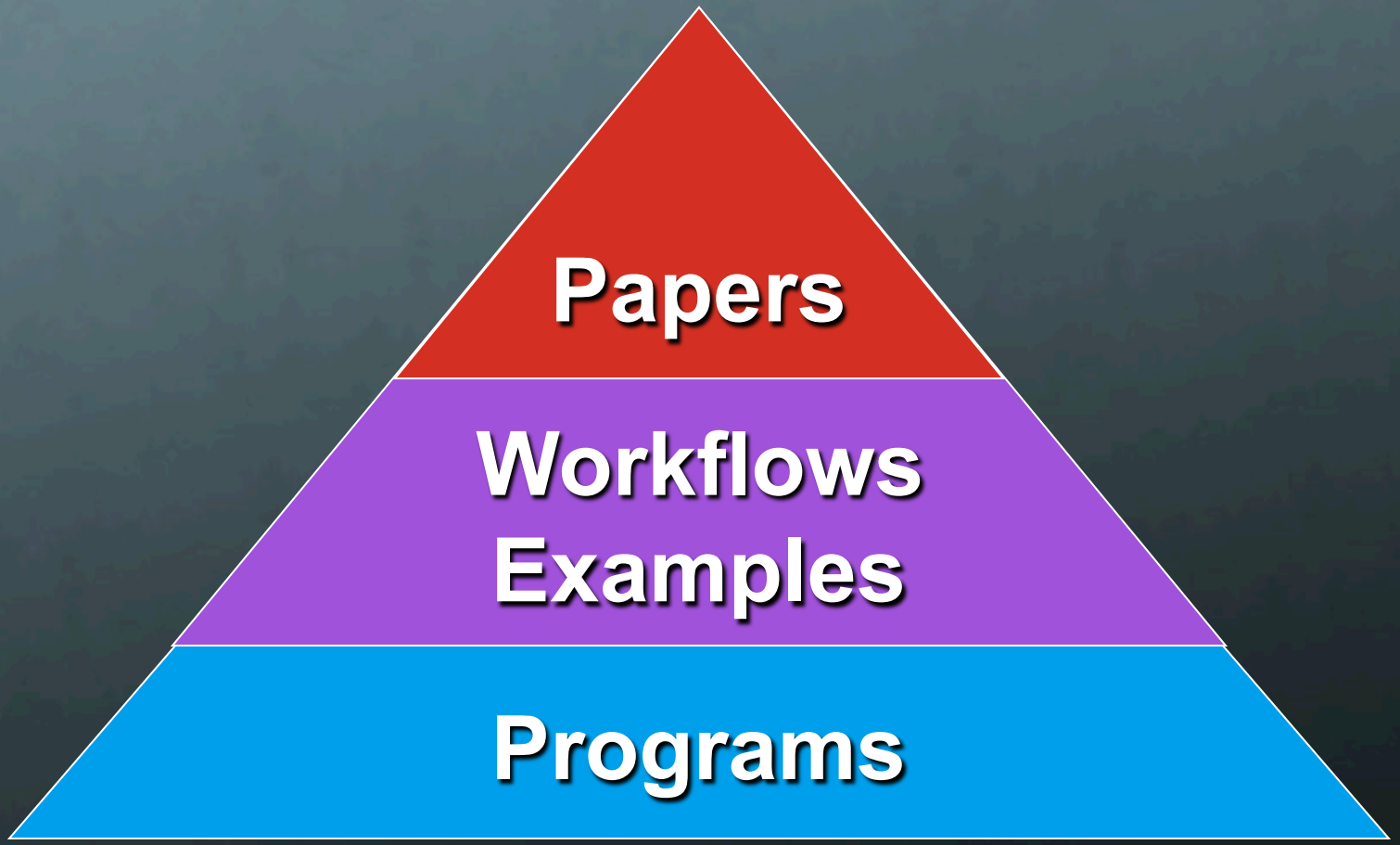
# ohloh.net about Madagascar

In a Nutshell, Madagascar…
… has had 9,158 commits made by 63 contributors
representing 556,787 lines of code
… is mostly written in C
with an average number of source code comments
… has a well established, mature codebase
maintained by a very large development team
with increasing Y-O-Y commits
… took an estimated 149 years of effort (COCOMO)
starting with its first commit in May, 2003
ending with its most recent commit 4 days ago

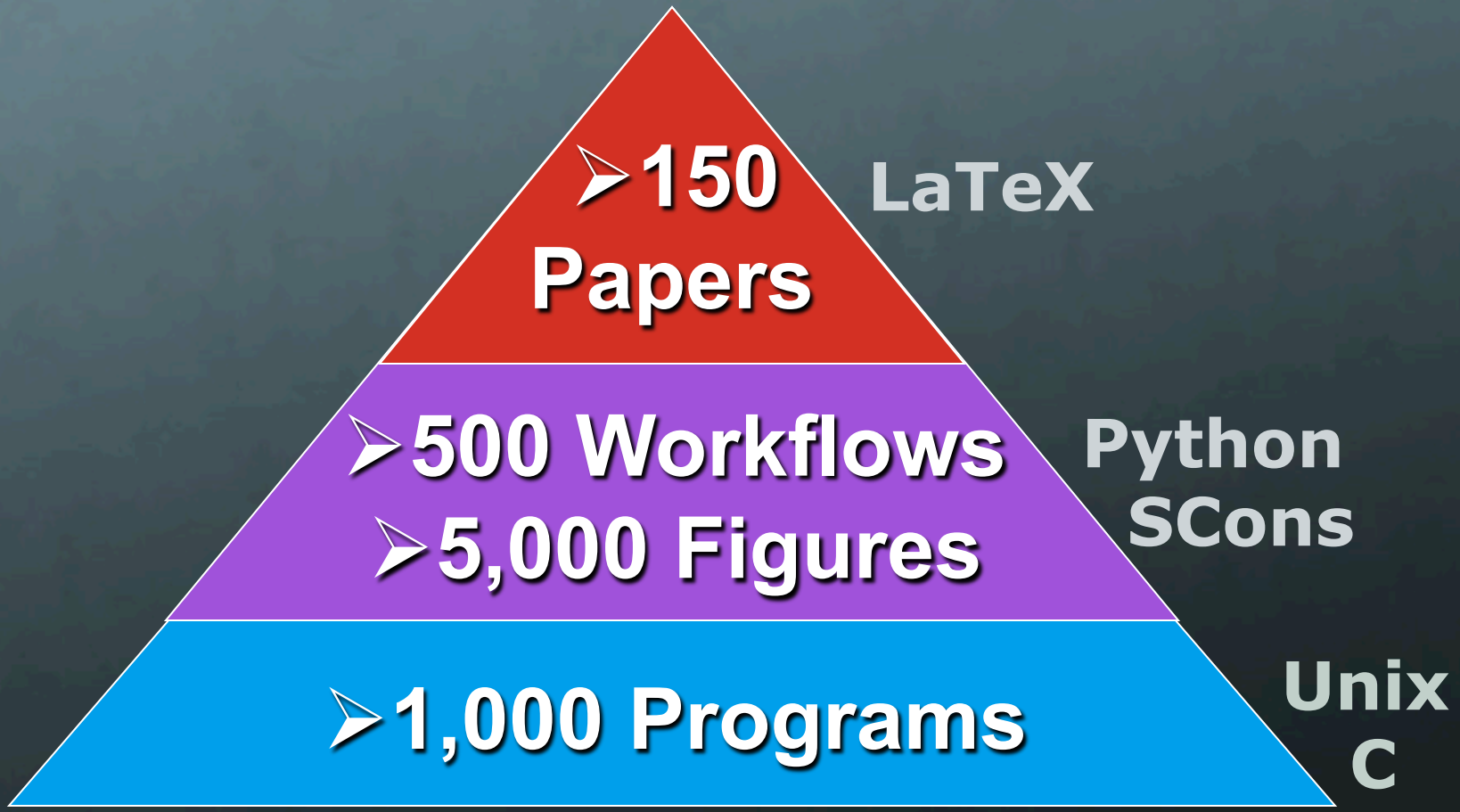
# ohloh.net about Madagascar



# Research Pyramid



# Research Pyramid



# Madagascar Workflow

- How does a user install my tool?
  - Python framework
- How does my tool capture experimental steps?
  - Editing SConstruct files
- How can documents link to the experiments?
  - Customized LaTeXHTML links
- Does your tool provide support for archival and longevity?
  - Community effort + regression testing

http://www.ahay.org/