Using Docker Containers with a Common Graphics User Interface to Address the Reproducibility of Research

DANIEL KRISTIYANTO
MS - COMPUTER SCIENCE
@kristiyanoto
Motivation

The cancer test, Jocelyn Kaiser


6 of 53 Cancer papers that Amgen could reproduce

14 of 67 Biomedical papers that Bayer completely reproduced

55% MD Anderson researchers who could not reproduce a published study
Challenges

Computational Biology:

- Involve complicated pipelines with many tools, multiple OS’es
- Software versioning
- Dependency and configuration constraints
Our Solution

We distribute the entire computing environment—not just the codes.
Virtual Machines

[ OSes running inside OS ]

- Bundle everything as a single machine

Docker Containers

Similar to VM, but with additional advantages:

- Smaller, easy to distribute
- Straightforward pipelines
- Easier to deploy
- Open source
- Build on top of other containers
Proof of concept: **Gene Network Inference**

**Maciej Fronczuk** | UWTacoma

*Source Code for Biology and Medicine 2015 10:11*
DOI: 10.1186/s13029-015-0043-5

**Chad Young** | UWSeattle

GUIdock: Using Docker Containers with a Common Graphics User Interface to Address the Reproducibility of Research

Ling-Hong Hung*, Daniel Kristiyanto*, Sung Bong Lee*, Ka Yee Yeung*
Institute of Technology, University of Washington, Tacoma, WA 98402, United States of America

* These authors contributed equally to this work.
† kayee@uw.edu

Paper: http://dx.doi.org/10.1371/journal.pone.0152686
Repository: http://github.com/biodepot
Acknowledgement

Dr. Ling-Hong Hung, Sung Lee, Dr. Ka Yee Yeung
noVNC:

- Works better in the cloud
- More robust graphic and for more complex pipelines
Further Works

Trevor Meiss | UWT

DToxS: Reproducibility of RNAseq analyses (in collaboration with Mount Sinai, NYC)

DNA Alignment and differentially expressed genes identification
BioDepot

http://tacoma.uw.edu/bioinformatics

Repository of containerized bioinformatics pipelines