

Solutions and technologies for promoting reproducibility

A developer's perspective

Dan Leehr
Center for Genomic and Computational Biology
Duke TechExpo 2015

Research Needs

- Word processor
- Spreadsheet/statistics software
- Dissemination of results
- Reproducible/repeatable methods for running analysis
- Process support (Experimentation, Validation)
- Management of hardware and software
- Notebook for results

Software Development Tools

- Version Control and Code hosting (Git, GitHub)
- Plaintext / Markdown
- IDEs/Literate Programming (RStudio, IPy Notebook)
- Build tools (scripts, Make)
- Continuous Integration (Travis CI, Jenkins)
- Virtualization/Containerization (VM, Docker)

Software Development Tools

- Git for more than code, gh-pages
- Markdown, RMarkdown, RST
- RMd + KnitR: Words + Code + Data = Paper!
- Build tools to generate research products rather than programs
- CI/CD to automate above steps
- Containers for reproducible environments

What else?