Reproducibility for Bioinformatic tools and workflows
Authors: Peter Crisp, UMN, USA (0000-0002-6327-0113); Deep Ganguly, ANU, AUS (0000-0001-5746-0141)

Documentation

Why did I do this?

How can I record the steps of a bioinformatic analysis?

Version Control

What version of the program, data etc... did I use?

Version conflict

Software

Dependency hell

28% of the tools are impossible to install

Git

Records and illustrates changes between versions

Git tracks changes to files over time, enabling documentation of development and reversion to prior versions.

Managers

Managers and containers enable consistent use of software when emulating or reproducing an analysis.

Containers

Docker runs images as containers that are self contained with all code, programs, libraries included. No subsequent installation required.

Conda, Bioconda

Conda and Bioconda are tools for managing software dependencies across different platforms.

Managers

Managers allow for multiple independent environments

Software

28% of the tools are impossible to install

Notebooks

Notebooks are an essential tool for documenting analyses, enabling reproducibility and sharing.

Git

Git tracks changes to files over time, enabling documentation of development and reversion to prior versions.

Resources & further reading

Reproducible computational research
Sandve et al. (2013), PLoS Comp Bio, https://doi.org/10.1371/journal.pcbi.1003285

Software: reproducible installation

Documentation

‘Bioinformatic data skills’ Buffalo
http://shop.oreilly.com/prod/0636920030157.do
Software carpentry https://software-carpentry.org/

Version control:
http://smutch.github.io/VersionControlTutor/

Containers:
https://docs.docker.com/
Biocontainers http://biocontainers.pro

Reviewing:
Computation checklist github.com/vivekbhr/reproChecklist

Resources →

Topic →

Challenges →

Solutions →

Reproducible computational research
Sandve et al. (2013), PLoS Comp Bio, https://doi.org/10.1371/journal.pcbi.1003285

Software: reproducible installation

Documentation

‘Bioinformatic data skills’ Buffalo
http://shop.oreilly.com/prod/0636920030157.do
Software carpentry https://software-carpentry.org/

Version control:
http://smutch.github.io/VersionControlTutor/

Containers:
https://docs.docker.com/
Biocontainers http://biocontainers.pro

Reviewing:
Computation checklist github.com/vivekbhr/reproChecklist

Challenges →

Solutions →

Resources →

Reproducible computational research
Sandve et al. (2013), PLoS Comp Bio, https://doi.org/10.1371/journal.pcbi.1003285

Software: reproducible installation

Documentation

‘Bioinformatic data skills’ Buffalo
http://shop.oreilly.com/prod/0636920030157.do
Software carpentry https://software-carpentry.org/

Version control:
http://smutch.github.io/VersionControlTutor/

Containers:
https://docs.docker.com/
Biocontainers http://biocontainers.pro

Reviewing:
Computation checklist github.com/vivekbhr/reproChecklist

Reproducible computational research
Sandve et al. (2013), PLoS Comp Bio, https://doi.org/10.1371/journal.pcbi.1003285

Software: reproducible installation

Documentation

‘Bioinformatic data skills’ Buffalo
http://shop.oreilly.com/prod/0636920030157.do
Software carpentry https://software-carpentry.org/

Version control:
http://smutch.github.io/VersionControlTutor/

Containers:
https://docs.docker.com/
Biocontainers http://biocontainers.pro

Reviewing:
Computation checklist github.com/vivekbhr/reproChecklist