

# **DEVOPS TO RESOPS: A FRAMEWORK OF COMMUNITY BEST PRACTICES AND TOOLS TO GUIDE COMPUTATIONAL REPRODUCIBILITY**

**AARON WILLCOX**

[@aaron\\_willcox](#) [twitter](#) [linkedin](#)  
**2022-07-07**

# RESEARCH DATA & CODE

- Source code generated each year grows by about 20% (L. Hatton & M. van Genuchten, 2019).
- 93% of authors fail to respond or declined sharing data.
- PLOS journals data sharing <30%
- Sharing policy increase: 15% in 2015 to 75% in 2020 (Culina et al., 2018).
- Data handling and processing often informally transmitted (Maer-Matei et al., 2019).
- Lack of formal training for researchers (Koehler Leman et al., 2020).

# COMPUTATIONAL REPRODUCIBILITY

*The ability to produce equivalent analytical outcomes from the same data set using the same **code** and **software** as the original study (Fidler et al., 2017).*

- Why is there this gap between creation and sharing of knowledge?

# LETS BEGIN!

# SOURCES OF IRREPRODUCIBILITY

- Workflows
- Environment(Software dependencies)
- Version Control System
- Documentation (Leipzig et al., 2020)

**WHAT IF I TOLD YOU...**

# **PRACTICES DIFFER VERY LITTLE TO SOFTWARE ENGINEERS**

# DEVOPS

- Culture or working philosophy that fosters cross-functional team communication and collaboration.
- Orients the **way** in which a team collaborates.
- DevOps Centers on a culture of collaboration.
- ResOps should\* center on a culture of reproducibility.



# THE MODERN SCIENTIST ROLE

- Developers
- Testers
- Programmers
- Operators
- **One Role**

# **WE CREATE SOFTWARE**

# FAIR PRINCIPLES

- Findable
- Accessible
- Interoperable
- Reusable
- **Not a technical solution or standard**

# VERSION CONTROL

- Analysis\_Report\_2022\_v001\_final\_FINAL\_Submitted.pdf

# VERSION CONTROL

- Git command line and communication protocol
- Github Web platform
- Re-purposed for research

# CONTINUOUS PRACTICES

- **Continuous Integration:** Integrate code, automate, test
- **Continuous Delivery:** testing, dynamic documents
- **Continuous Deployment:** packaging and releases

# INCREASE YOUR DOCUMENTATION GAME

- Document text code and data
- README and Wikis
- Annotating Code
- YAMLS
- Vignettes

# CONTAINER TECHNOLOGY (DOCKER)

- Lock and isolate the environment
- Manage Dependencies
- Isolate Analysis from Local
- Reusable and Portable



# DOCKERFILE EXAMPLE

```
1 FROM rocker/verse:4.2.0
2
3 LABEL version="1.0"
4 LABEL maintainer="aaron.willcox@unimelb.edu.au"
5 LABEL description="This is the Dockerfile \
6 for the replicATs pipeline including Rstudio, Tidyverse \
7 package checking and python support."
8
9 # installers
10 RUN apt-get -qq update && \
11     DEBIAN_FRONTEND=noninteractive apt-get -qq install -f \
12     # Translator for geospatial data
13     gdal-bin \
14     # compress large file in blocks
15     lbzip2 \
16     # scientific library
17     gsl-bin \
18     # Gibbs samplers
19     jags \
20     r-cran-rjags \
21     # terminal editor
22     nano \
23     # common data formatting
24     netcdf-bin \
25     && apt-get clean
26
27 # install python libraries
28 RUN /rocker_scripts/install_python.sh
29
30 # Additional libraries
31 RUN install2.r --error --skipinstalled \
32     cli \
33     osfr \
34     pointblank \
35     qualtrics \
36     rjags \
37     R2jags \
38     && rm -rf /tmp/downloaded_packages/ /tmp/*.rds
39
40 # Install packages from github
41 RUN R -e "remotes::install_github('jthomasmock/gtExtras')"
42
```

```
43 # Install Pins version
44 RUN R -e "remotes::install_version('pins', version = '0.4.5')"
45
37:11
```

Dockerfile : |

# NEW PARADIGM

- **Scientific Workflows**
- Version Control <| |> Documented History
- Docker <| |>| System Environment & Dependencies
- CI/CD <| |> Testing, delivery, Rapid feedback
- *“Power in incorporating them into a scientific workflow greatly increases the ability for its reusability”*

# RESOPS

*ResOps is an attempt to distinguish scientific software from DevOps. Emphasising communities of practice that enable the convergence of tools implemented in the scientific community to enable computational reproducibility in the form of a scientific workflow.*

# INDUNDATED WITH CHOICE?

- Look to the scientific community.
- Not a technical problem but a educational problem.

# PUBLISH YOUR CODE! IT IS GOOD ENOUGH!

*Wants to share a trade secret with scientists:  
Most professional computer software isn't very good. That the code is a little raw is one of the main reasons scientists give for not sharing it with others. So if your code is good enough to do the job then it is good enough to release  
(Barnes, 2010).*

# **SEND IT TO THE INTERNET**



# THANK YOU!

*@aaron\_willcox* 

# References

- AlNoamany, Yasmin, and John A. Borghi. 2018. "Towards Computational Reproducibility: Researcher Perspectives on the Use and Sharing of Software." Article. *PEERJ COMPUTER SCIENCE*, September. <https://doi.org/ghkb9j> (<https://doi.org/ghkb9j>).
- Archmiller, Althea A., Andrew D. Johnson, Jane Nolan, Margaret Edwards, Lisa H. Elliott, Jake M. Ferguson, Fabiola Iannarilli, et al. 2020. "Computational Reproducibility in the Wildlife Society's Flagship Journals." Article. *JOURNAL OF WILDLIFE MANAGEMENT* 84 (5): 1012–17. <https://doi.org/gg66q7> (<https://doi.org/gg66q7>).
- Beaulieu-Jones, Brett K, and Casey S Greene. 2017. "Reproducibility of Computational Workflows Is Automated Using Continuous Analysis." *Nature Biotechnology*, no. 4: 342.
- Catlin, Ann Christine, Chandima Hewa Nadungodage, and Andres Bejarano. 2019. "Lifecycle Support for Scientific Investigations: Integrating Data, Computing, and Workflows." Article. *COMPUTING IN SCIENCE & ENGINEERING* 21 (4): 49–61. <https://doi.org/gjt725> (<https://doi.org/gjt725>).
- Culina, Antica et al. 2018. "Navigating the Unfolding Open Data Landscape in Ecology and Evolution." *Nature Ecology & Evolution*, 1–7.
- de Bayser, Maximilien, Leonardo G. Azevedo, and Renato Cerqueira. 2015. "ResearchOps: The Case for DevOps in Scientific Applications." In *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, 1398–1404. Ottawa, ON, Canada: IEEE. <https://doi.org/ghkb9s> (<https://doi.org/ghkb9s>).
- Díaz, Jessica, Jorge Pérez-Martínez, Agustin Yague, Andrea Villegas, and Antonio Antona. 2019. *DevOps in Practice – A Preliminary Analysis of Two Multinational Companies*.
- Hatton, L., and M. Van Genuchten. 2019. "Computational Reproducibility: The Elephant in the Room." *IEEE Software* 36 (2): 137–44. <https://doi.org/ggkvtr> (<https://doi.org/ggkvtr>).
- Hocquet, Alexandre, and Frederic Wieber. 2021a. "Epistemic Issues in Computational Reproducibility: Software as the Elephant in the Room." Article. *EUROPEAN JOURNAL FOR PHILOSOPHY OF SCIENCE* 11 (2). <https://doi.org/gkm94m> (<https://doi.org/gkm94m>).
- Hocquet, Alexandre, and Frédéric Wieber. 2021b. "Epistemic Issues in Computational Reproducibility: Software as the Elephant in the Room." *European Journal for Philosophy of Science* 11 (2). <https://doi.org/gkm94m> (<https://doi.org/gkm94m>).
- Koehler Leman, Julia, Brian D. Weitzner, P. Douglas Renfrew, Steven M. Lewis, Rocco Moretti, Andrew M. Watkins, Vikram Khipple Mulligan, et al. 2020. "Better Together: Elements of Successful Scientific Software Development in a Distributed Collaborative Community." *PLoS Computational Biology* 16 (5): e1007507. <https://doi.org/ggt62r> (<https://doi.org/ggt62r>).
- Leipzig, Jeremy, Daniel Nüst, Charles Tapley Hoyt, Stian Soiland-Reyes, Karthik Ram, and Jane Greenberg. 2020. "The Role of Metadata in Reproducible Computational Research."
- Maer-Matei, Monica Mihaela, Tiberiu Marian Georgescu, Cristina Mocanu, and Ana-Maria Zamfir. 2019/01/01/". "Skill Needs for Early Career Researchers." *Sustainability* 11 (10). <https://doi.org/ghkb9m> (<https://doi.org/ghkb9m>).
- Murray, NJ et al. 2021. "Data Freshness in Ecology and Conservation." *Trends Ecol Evol* 36 (6): 485–87.
- Obels, Pepijn, Daniël Lakens, Nicholas Coles, Jaroslav Gottfried, and Seth Green. 2020. "Analysis of Open Data and Computational Reproducibility in Registered Reports in Psychology." *Advances in Methods and Practices in Psychological Science* 3 (May): 251524592091887. <https://doi.org/gg4vw4> (<https://doi.org/gg4vw4>).
- Peikert, Aaron, and Andreas Markus Brandmaier. 2019. "A Reproducible Data Analysis Workflow with R Markdown, Git, Make, and Docker." *DevArXiv*. <https://doi.org/10.31234/osf.io/8vzav> (<https://doi.org/10.31234/osf.io/8vzav>).