Reproducible workflows and collaboration with domain-knowledge experts

Charles T. Gray[1], Hollie Birkinshaw[2], Matthew Grainger[3], Charles A. Robinson[4], Tamar Pincus[2], and Gavin Stewart[1]

[1]Newcastle University
[2]Royal Holloway
[3]NINA
[4]CA

July 19, 2021

1 Unboxing the black box of data wrangling

Proponents of open science commendably underscore reuse and extensibility of scientific research components, such as data and code; techniques that facilitate the incorporation of these components into future analyses [1, 2, 3]. Less explicit attention, however, is given to the benefits of reproducible workflows, where results can be readily calculated by another researcher, and computational transparency *during* the research project, as opposed to beyond publication. Sure, we can appeal to an analyst's commitment to scientific civic duty, but by explicitly examining how reproducible workflows can facilitate collaboration with domain-knowledge experts we begin to answer the arguably more pertinent question, *What's in reproducibility for me*?

1.1 Benefits of reproducibility and open science

1.2 The problem of black box analysis: pitfalls, foibles, and outright fraud

1.3 A reproducible workflow for collaboration

2 Labelling scales and interpreting dosage

3 Discussion

References

[1] Heidi Laine. Open science and codes of conduct on research integrity. 37(4).

[2] R. D. Peng. Reproducible Research in Computational Science. 334(6060):1226–1227.

[3] Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. The FAIR Guiding Principles for scientific data management and stewardship. 3(1):160018.