Bibliographic Notes

Contents

1 FROM THE CS TYPES

What is Reproducibility? The R* Brouhaha [Gob16]. This is a presentation given by Goble at the Alan Turing Institute Symposium Reproducibility, Sustainability and Preservation , 6-7 April 2016, Oxford, UK. In this presentation the following terminology is used:

- **rerun**: variations of experiment and set up (*robust*)
- **repeat**: same experiment, same set up, same lab (*defend*)
- **replicate**: same experiment, same set up, independent lab (*certify*)
- **reproduce**: variations on experiment, on set up, independent labs (*compare*)
- **reuse**: different experiment (*transfer*)

Also distinguishes between "micro reproducibility" (within the research environment) and "macro reproducibility" (through peer review / publication).

Different uses: "personal productivity" and "reproducibility: public good" (cf. *provenance for self* vs *provenance for others'*).

Also mentions

- *levels of reproducibility* (portability vs depth)
- *repeatable environments* and *research objects*
- link to *FAIR*

PRIMAD [RBD+16]. Starting point for the development of the initial PRIMAD model [RBD+16] was a question raised[1] at the Dagstuhl seminar [FFR16]: "What is the *information gain* of a reproducibility study?" BL suggested that successful replication is often less informative than a failed one. But then again, it depends on the "delta" between the original study and the replication: e.g., if you keep "everything" the same (data, params, software, platform, experimenter, etc.) a successful replication provides little new information[2]. However, in this case, a failed replication consitutes a huge information gain: we must be missing something! Conversely, if we "wiggle" a lot (we have a large Δ) and change some or all of data, params, implementation, platform, human agents, etc.) and still get the "same" results, we learned a lot: the method/results/findings are robust! As a rule of thumb:

- the more elements you wiggle (or "prime" in PRIMAD parlance) and still get a positive outcome, the more robust the method (or claims, finding, etc.) and
- the fewer elements you wiggle and get a negative outcome, the more you learn about what you're missing!

Although PRIMAD is a first attempt to go beyond terminological clarification and will likely evolve, it has already been applied and studied [FFJ+16, GL17, CRB+19]. Hopefully more is coming ...

Provenance. [PFMB19]

2 TERMINOLOGIES

Sorting out Terminologies. Plesser [Ple18] tries to sort out the different terminologies and has a table that shows how Claerbout [CK92, SKC00] and ACM [ACM18] disagree and how they align with Goodman [GFI16]. Plesser states that "[GFI16] propose a new lexicon for research reproducibility with the following definitions: **Methods reproducibility**: provide sufficient detail about procedures and data so that the same procedures could be exactly repeated; **Results reproducibility** obtain the same results from an independent study with procedures as closely matched to the original study as possible; **Inferential reproducibility**: draw the same conclusions from either an independent replication of a study or a reanalysis of the original study." [Ple18] also praises [GFI16] as "an important step out of the terminology quagmire in which the active and fruitful debate about the trustworthiness of research has been stuck for the past decade, because it sidesteps confounding common language associations of terms by explicit labeling (explicit is better than implicit; Peters, 2004). One can only wish that it will be adopted widely so that the debate can once more focus on scientific rather than language issues."

3 FROM THE SCIENCES

Replicability or reproducibility? On the replication crisis in computational neuroscience and sharing only relevant detail [MHH18]. From the abstract: "we draw on methodological studies into the replicability of psychological experiments and on the mechanistic account of explanation to analyze the functions of model replications and model reproductions in computational neuroscience. We contend that **model replicability**, or independent researchers' ability to obtain the same output using original code and data, and **model reproducibility**, or independent researchers' ability to recreate a model without original code, **serve different functions and fail for different reasons**. This means that **measures designed to improve model replicability may not enhance (and, in some cases, may actually damage) model reproducibility**. We claim that although both are undesirable, low model reproducibility poses more of a threat to long-term scientific progress than low model replicability. In our opinion, low model reproducibility stems mostly from authors' **omitting to provide crucial information in scientific papers and we stress that**

[1] by Bertram :-)
[2] other than the algorithm/method seems to be deterministic!

sharing all computer code and data is not a solution. Reports of computational studies should remain selective and include all and only relevant bits of code."

So this makes a big point about **model replication** vs. **model reproduction**. Here's another excerpt:

> Only by **reproducing** a model, or offering what **psychologists** would call a **conceptual replication**, can we discover the model's hidden assumptions, bugs or unexpected interactions. Moreover, **a successful reproduction contributes to model validation**. **Validation** is methodologically more valuable than **verification** because it shows **how an implemented model corresponds to empirical data**. In particular, by validating a computational model through reproduction we make sure that the results of modeling are sound (see Drummond 2009 for a similar argument).

More quotes:

> As far as model reproducibility is concerned, specific rules such as those introduced by Sandve et al. (2013) are of **little help** unless they contribute to an understanding of how theoretical principles are translated into modeling practice. **Indeed, they can improve replicability at the cost of affecting reproducibility**. Specifying the **exact computational platform should not be required as long as software is designed to be portable for interoperability.** With the exception of the use of supercomputers, specialized hardware or very sensitive real- time requirements, if the details of the computational architecture count then the model may be too sensitive to improper background conditions, and its scientific value is limited. It is simply poor-quality code. Thus, while avoiding prototype code and quick hacks is difficult in actual scientific work, **this kind of code should be discardable**. It should be possible to reconstruct the model with no knowledge of such detail, from a paper only. What we suggest, then, is that **a paper should above all contain all and only information needed to reproduce a model and assess the model's intended relationship to its target.** By contrast, information necessary to **replicate** a model, including code and experimental data sets, should be deposited in open repositories (Migliore et al. 2003). Research papers and repositories serve different functions, though ultimately both should contribute to the same overarching goal of making science cumulative. If science communication continues to rely on journal papers then too much irrelevant detail will make reading (and reviewing) them impossible. Open model repositories facilitate replication and code reuse.

Open is not enough [CDD+19]. Subtitle / tag-line : "The solutions adopted by the high-energy physics community to foster reproducible research are examples of best practices that could be embraced more widely. **This first experience suggests that reproducibility requires going beyond openness**." Includes table corresponding to the"6 Rs" from Goble [Gob16] (also credits Barba); then goes on to refine that model for HEP context. Mentions *guiding principles towards reproducibility*:

- Define your reproducibility goals
- Incorporate best practices early in your research
- Build on what is there
- Structure your knowledge
- Capture your workflows
- Raise awareness
- Embrace openness whenever possible
- Enable liberal and fair reuse

Mentions in the conclusions: "Sharing data is not enough; it is also **essential to capture the structured information about the research data analysis workflows and processes** to ensure the usability and longevity of results."

REFERENCES

[ACM18] ACM. Artifact Review and Badging, April 2018. https://www.acm.org/publications/policies/artifact-review-badging.

[CDD+19] X. Chen, S. Dallmeier-Tiessen, R. Dasler, S. Feger, P. Fokianos, J. B. Gonzalez, H. Hirvonsalo, D. Kousidis, A. Lavasa, S. Mele, D. R. Rodriguez, T. Šimko, T. Smith, A. Trisovic, A. Trzcinska, I. Tsanaktsidis, M. Zimmermann, K. Cranmer, L. Heinrich, G. Watts, M. Hildreth, L. L. Iglesias, K. Lassila-Perini, and S. Neubert. Open Is Not Enough. *Nature Physics*, 15(2):113, February 2019.

[CK92] J. F. Claerbout and M. Karrenbach. Electronic Documents Give Reproducible Research a New Meaning. In *SEG Technical Program Expanded Abstracts 1992*, pp. 601–604. Society of Exploration Geophysicists, 1992.

[CRB+19] D. Chapp, D. Rorabaugh, D. Brown, E. Deelman, K. Vahi, V. Welch, and M. Taufer. Applicability Study of the PRIMAD Model to LIGO Gravitational Wave Search Workflows. *arXiv:1904.05211 [astro-ph]*, April 2019.

[FFJ+16] N. Ferro, N. Fuhr, K. Järvelin, N. Kando, M. Lippold, and J. Zobel. Increasing Reproducibility in IR: Findings from the Dagstuhl Seminar on "Reproducibility of Data-Oriented Experiments in e-Science". *SIGIR Forum*, 50(1):68–82, June 2016.

[FFR16] J. Freire, N. Fuhr, and A. Rauber. Reproducibility of Data-Oriented Experiments in e-Science (Dagstuhl Seminar 16041). *Dagstuhl Reports*, 6(1):108–159, 2016.

[GFI16] S. N. Goodman, D. Fanelli, and J. P. A. Ioannidis. What Does Research Reproducibility Mean? *Science Translational Medicine*, 8(341), June 2016.

[GL17] M. R. Gryk and B. Ludäscher. Workflows and Provenance: Toward Information Science Solutions for the Natural Sciences. *Library Trends*, 65(4):555–562, September 2017.

[Gob16] C. Goble. What Is Reproducibility? The R* Brouhaha, April 2016.

[MHH18] M. Miłkowski, W. M. Hensel, and M. Hohol. Replicability or Reproducibility? On the Replication Crisis in Computational Neuroscience and Sharing Only Relevant Detail. *Journal of Computational Neuroscience*, 45(3):163–172, December 2018.

[PFMB19] J. F. Pimentel, J. Freire, L. Murta, and V. Braganholo. A Survey on Collecting, Managing, and Analyzing Provenance from Scripts. *ACM Comput. Surv.*, 52(3):47:1–47:38, June 2019.

[Ple18] H. E. Plesser. Reproducibility vs. Replicability: A Brief History of a Confused Terminology. *Frontiers in Neuroinformatics*, 11, 2018.

[RBD+16] A. Rauber, V. Braganholo, J. Dittrich, N. Ferro, J. Freire, N. Fuhr, D. Garijo, C. Goble, K. Järvelin, B. Ludäscher, B. Stein, and R. Stotzka. PRIMAD: Information Gained by Different Types of Reproducibility. In Freire et al. [FFR16], pp. 128–132.

[SKC00] M. Schwab, N. Karrenbach, and J. Claerbout. Making Scientific Computations Reproducible. *Computing in Science Engineering*, 2(6):61–67, November 2000.