The Dagstuhl Beginners Guide to Reproducibility for Experimental Networking Research

SIGCOMM 2019
Beijing, China

SIGCOMM Computer Communication Review
January 2019 (Editorial)
https://doi.org/10.1145/3314212.3314217

- Vaibhav Bajpai *Technische Universität München*
- Anna Brunstrom *Karlstad University*
- Anja Feldmann *MPI für Informatik*
- Wolfgang Kellerer *Technische Universität München*
- Aiko Pras *University of Twente*
- Henning Schulzrinne *Columbia University*
- Georgios Smaragdakis *TU Berlin*
- Matthias Wählisch *Freie Universität Berlin*
- Klaus Wehrle *RWTH Aachen*

August 21, 2019

Motivation

- Reproducibility is the cornerstone of the scientific process.
- *Yet,* lack of reproducibility exists an ongoing problem. For instance:

  A survey [1] of MANET simulation studies (2000-2005) found only 15% papers were repeatable.

  A study [2] (2009) explored 134 TOIP papers and found few release code (9%) and data (33%).

  A study [3] (2016) examined 601 ACM papers and found only 32% to be repeatable.

We believe,

- There is a need to inculcate the importance of reproducibility at an early-stage.
- A beginners guide that documents current best practises helps students *embrace* reproducibility.

Terminology

ACM Terminology [4]

- Repeatability. *same team, same experimental setup.*
- Replicability. *different team, same experimental setup.*
- Reproducibility. *different team, different experimental setup.*

should (ideally) only require general knowledge of the discipline + paper + artefacts.

Goals and Principles

- supports continuation and building on earlier work of own and others.
- avoids reverse-engineering previously written code.
- increases trust in experimental data gathered by own and others.
- reduces likelyhood of making mistakes (or at least easier to find).

Best Practises

Best Practises

- Problem Formulation and Design
- Documentation
- Experimentation and Data Collection
- Handling Data

Best Practises | Problem Formulation and Design

Hypothesize. *think first, run later.*

- Formulate hypothesis → design → conduct experiment → check the hypothesis.
- Double check results to spot errors (with advisor, teammates)

Plan and solicit early feedback

- Visualisations help explain results and spot anomalies (notches, spikes, gaps).
- Explore the parameter space (ANOVA). Get feedback often.

Iterate

- Record steps and automate them (scripts, Makefiles).
- Account for factors (time of day) that may affect one-time measurments.

Factor dynamism

- Expect that operational systems would not remain static during experimentation.

Best Practises | Documentation

Record the experiment

- Use lab notebooks. Record all steps and observations (mistakes too).
- Avoid temptation to skip documentating code for later. Research artefacts are reused.

Treat metadata as data

- How data was created, what it contains, where it's documented, how to recreate it.

Use a version control system

- VCS helps identify source of change in measured results.
- Create publishable results by creating release of your software.

Keep regular backups

- Data management plans for research grants require artefacts to be preserved for years.

Best Practises | Experimentation and Data Collection

Validate and scale. *start small, then expand.*

- Starting small helps readily predict results and verify tools.
- Use test-cases as sanity during regression and scaling up of components.

Do not reinvent the wheel. *do one thing, and one thing well.*

- Check whether the tool that solves the problem at hand, already exists.
- Creating your own tool, also commits you into maintaining it.

Monitor your experiment

- Monitor your operational system to avoid common problems:
  disk out of space, machine reboots, overwritten logs, wrong permissions, network failures.

Best Practises | Handling Data

Data privacy, data anonymization and ethics

- Never try to de-anonymize data (unethical, discourages others from making data available)
- Think about privacy concerns when releasing data (consider anonymization)
- Seek consultation (team members, seniors, ethics panels, IRB) when in doubt.
- Refer to published community ethics guidelines [5, 6]

Data integrity. *account for observation bias.*

- Evaluate the performance complexity of the system based on its intended use-case.

Licensing and giving credit

- Consult with everyone in the team to agree on how code intends to be licensed.:
- Some licenses require modifications to be made publicly available.
- Some licenses [7, 8] mandate giving credit to sources

Further Reading | What should be Documented?

Guidelines for specific research methodologies:

- Simulations
- Systems Prototyping and Evaluations
- Human Subject and Subjective Experiments
- Real-world Measurements

Please refer to the paper [9] for details

A must read for graduate students before starting on a related project!

State of Reproducibility

Past, Present, and Future

State of Reproducibility | Reproducibility Course and SIGCOMM Workshops

2012 Stanford's reproducibility course.

https://reproducingnetworkresearch.wordpress.com

2017 CCR article reporting past 5 years of experience from running the course [10]

- 200 students, 40 networking papers, 3 weeks duration, working in pairs

Learning Networking by Reproducing Research Results

Lisa Yan
Stanford University
yanlisa@stanford.edu

Nick McKeown
Stanford University
nickm@stanford.edu

2017 SIGCOMM Workshop on Reproducibility [11] (a related workshop was held in 2003 [12])

Thoughts and Recommendations from the ACM SIGCOMM 2017 Reproducibility Workshop

Damien Saucez
Univresité Côte d'Azur, Inria, France
damien.saucez@inria.fr

Luigi Iannone
Telecom Paristech, France
luigi.iannone@telecom-paristech.fr

State of Reproducibility | Artefacts Evaluation and Reproducibility Track

2017 CCR article on artefacts availability in accepted papers [10]

- SIGCOMM, CoNEXT, IMC, ICN conferences
- 49/137 responses from authors, 35.8%
- Webpage: https://artefacts.cm.in.tum.de/2017

2018 SIGCOMM Artifacts Evaluation Committee (AEC) [13].

- 32 accepted papers were submitted, 28 were badged.

2018 CoNEXT badged accepted papers (will be continued in 2019).

- 14/32 accepted papers submitted for evaluation, 12 papers badged.

2019 IMC reproducibility track [14] solicits work that reproduces previous work.

2018 Dagstuhl seminar #18412 [15] on Encouraging Reproducibility in Scientific Internet Research

- New publication strategies [16]
- Incentives and ontology for reproducibility
- Reproducibility in post-publication phase
- Reproducibility track for IMC
- Guidelines for students [9] and reviewers [17]

The Dagstuhl Beginners Guide to Reproducibility for Experimental Networking Research

Vaibhav Bajpai
TU Munich
bajpaiv@in.tum.de

Anna Brunstrom
Karlstad University
anna.brunstrom@kau.se

Anja Feldmann
MPI for Informatics
anja@mpi-inf.mpg.de

Wolfgang Kellerer
TU Munich
wolfgang.kellerer@tum.de

Aiko Pras
University of Twente
a.pras@utwente.nl

Henning Schulzrinne
Columbia University
hgs@cs.columbia.edu

Georgios Smaragdakis
TU Berlin
georgios@inet.tu-berlin.de

Matthias Wählisch
Freie Universität Berlin
m.waehlisch@fu-berlin.de

Klaus Wehrle
RWTH Aachen University
klaus@comsys.rwth-aachen.de

Report from Dagstuhl Seminar 18412

Encouraging Reproducibility in Scientific Research of the Internet

Edited by

Vaibhav Bajpai[1], Olivier Bonaventure[2], Kimberly Claffy[3], and Daniel Karrenberg[4]

Summary

- Best Practises
  - Problem Formulation and Design
  - Documentation
  - Experimentation and Data Collection
  - Handling Data
- Guidelines for Specific Methodologies
  - Simulations
  - Systems Prototyping and Evaluations
  - Human Subject and Subjective Experiments
  - Real-world Measurements

We hope the guide can serve as a key resource for graduate students and helps improve the state of reproducibility in experimental networking research.

www.vaibhavbajpai.com

bajpaiv@in.tum.de | @bajpaivaibhav

References

[1] S. Kurkowski, T. Camp, and M. Colagrosso, "MANET simulation studies: The incredibles," *Mobile Computing and Communications Review*, vol. 9, no. 4, pp. 50–61, 2005. [Online]. Available: http://doi.acm.org/10.1145/1096166.1096174

[2] P. Vandewalle, J. Kovacevic, and M. Vetterli, "Reproducible Research in Signal Processing," *IEEE Signal Processing Magazine*, vol. 26, no. 3, pp. 37–47, May 2009.

[3] C. S. Collberg and T. A. Proebsting, "Repeatability in computer systems research," *Communications of the ACM*, vol. 59, no. 3, pp. 62–69, 2016. [Online]. Available: http://doi.acm.org/10.1145/2812803

[4] ACM. (2016) Artifact review and badging. [Online]. Available: https://www.acm.org/publications/policies/artifact-review-badging

[5] David Dittrich and Erin Kenneally. (2012) The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research. [Online]. Available: https://www.dhs.gov/publication/csd-menlo-report

[6] Michael Bailey, David Dittrich, and Erin Kenneally. (2013) Applying Ethical Principles to Information and Communication Technology Re- search: A Companion to the Menlo Report. [Online]. Available: https://www.dhs.gov/publication/csd-menlo-companion

[7] Open Source Initiative. (2018) Licenses and Standards. [Online]. Available: https://opensource.org/licenses

[8] Creative commons. [Online]. Available: https://creativecommons.org

[9] V. Bajpai, A. Brunström, A. Feldmann, W. Kellerer, A. Pras, H. Schulzrinne, G. Smaragdakis, M. Wählisch, and K. Wehrle, "The Dagstuhl Beginners Guide to Reproducibility for Experimental Networking Research," *Computer Communication Review*, vol. 49, no. 1, pp. 24–30, 2019. [Online]. Available: https://doi.org/10.1145/3314212.3314217

[10] L. Yan and N. McKeown, "Learning networking by reproducing research results," *Computer Communication Review*, vol. 47, no. 2, pp. 19–26, 2017. [Online]. Available: https://doi.org/10.1145/3089262.3089266

[11] D. Saucez and L. Iannone, "Thoughts and Recommendations from the ACM SIGCOMM 2017 Reproducibility Workshop," *Computer Communication Review*, vol. 48, no. 1, pp. 70–74, 2018. [Online]. Available: https://doi.org/10.1145/3211852.3211863

[12] Workshop on Models, Methods and Tools for Reproducible Network Research (MoMeTools). [Online]. Available: https://conferences.sigcomm.org/sigcomm/2003/workshop/mometools

[13] D. Saucez, L. Iannone, and O. Bonaventure, "Evaluating the artifacts of SIGCOMM papers," *Computer Communication Review*, vol. 49, no. 2, pp. 44–47, 2019. [Online]. Available: https://doi.org/10.1145/3336937.3336944

[14] Reproducibility Track at IMC 2019. [Online]. Available: https://conferences.sigcomm.org/imc/2019/call-for-posters

[15] V. Bajpai, O. Bonaventure, K. C. Claffy, and D. Karrenberg, "Encouraging Reproducibility in Scientific Research of the Internet

(Dagstuhl Seminar 18412)," *Dagstuhl Reports*, vol. 8, no. 10, pp. 41–62, 2018. [Online]. Available: https://doi.org/10.4230/DagRep.8.10.41

[16] A. Dainotti, R. Holz, M. Kühlewind, A. Lutu, J. Sommers, and B. Trammell, "Open collaborative hyperpapers: a call to action," *Computer Communication Review*, vol. 49, no. 1, pp. 31–33, 2019. [Online]. Available: https://doi.org/10.1145/3314212.3314218

[17] D. K. D. S. Olivier Bonaventure, Luigi Iannone. (2018) ACM SIGCOMM Artefact Review Form. [Online]. Available: https://goo.gl/JjXgjw