ReprohackNL 2019: how libraries can promote research reproducibility through community engagement

Kristina Maria Hettne[1], Ricarda Proppert, Linda Nab, L. Paloma Rojas-Saunero, Daniela Gawehns

Abstract

University Libraries play a crucial role in moving towards Open Science, contributing to more transparent, reproducible and reusable research. The Center for Digital Scholarship (CDS) at Leiden University (LU) library is a scholarly lab that promotes open science literacy among Leiden's scholars by two complementary strategies: existing top-down structures are used to provide training and services, while bottom-up initiatives from the research community are actively supported by offering the CDS's expertise and facilities. An example of how bottom-up initiatives can blossom with the help of library structures such as the CDS is ReproHack. ReproHack – a reproducibility hackathon – is a grass-root initiative by young scholars with the goal of improving research reproducibility in three ways. First, hackathon attendees learn about reproducibility tools and challenges by reproducing published results and providing feedback to authors on their attempt. Second, authors can nominate their work and receive feedback on their reproducibility efforts. Third, the collaborative atmosphere helps building a community interested in making their own research reproducible.

A first ReproHack in the Netherlands took place on November 30th, 2019, co-organised by the CDS at the LU Library with 44 participants from the fields of psychology, engineering, biomedicine, and computer science. For 19 papers, 24 feedback forms were returned and five papers were reported as successfully reproduced. Besides the researchers' learning experience, the event led to recommendations on how to enhance research reproducibility. The ReproHack format therefore provides an opportunity for libraries to improve scientific reproducibility through community engagement.

Introduction

University Libraries around the world play a crucial role in open science, contributing to more transparent, reproducible and reusable research by offering training and services to their researchers. The Center for Digital Scholarship (CDS) at Leiden University (LU), The Netherlands, is a scholarly lab located in the LU Library. The CDS employs two complementary strategies to improve open data literacy among Leiden's scholars: existing top-down structures are used to provide 'training[2]', such as a brief training on data publishing, and a more extensive data FAIRification workshop, where researchers get hands on experience with making their data Findable, Accessible, Interoperable and Reusable (FAIR) (Wilkinson, et al., 2016) according to recent recommendations (Jacobsen, et al., 2019). Complementary to such training, bottom-up initiatives are actively supported by offering the CDS's expertise and facilities. Examples of such bottom-up initiatives initiated by Leiden's research community are data management workshops addressing institute-specific data challenges, for

example, 'pilot projects[3]' on making book history data FAIR and an ongoing initiative developing 'guidelines[4]' on how to share Medical Resonance Imaging data.

While many libraries make an effort to improve data access, few have focussed on reproducibility as another crucial aspect of open science literacy. Reproducibility can be defined as 'data *and code* being available to fully rerun the analysis' (The Turing Way Community, et al., 2019) and has become a widely discussed issue over the past decade (e.g., for an early discussion on the reproducibility crisis in psychological science, see (Pashler & Wagenmakers, 2012). In 2016, a Nature survey (Baker, 2016) listed factors contributing to irreproducible research. The top factors, such as intense competition and time pressure, can be tackled by scientific organizations adjusting their policies and incentive structures in a top-down approach. However, factors such as 'Methods, code unavailable', 'Raw data not available from original lab' and 'Insufficient peer review' may be more effectively tackled in a bottom-up fashion by enabling the scientific community in their practices, and provide a low hanging fruit for librarians to pick.

One initiative that addresses these factors is 'ReproHack[5]'. ReproHack is a grass-root initiative by young scholars with the goal of improving research reproducibility in three ways: First, hackathon attendees learn about reproducibility tools and challenges by reproducing published results and providing feedback to authors on their attempt. Second, authors can nominate their own work and receive feedback on their reproducibility efforts. Third, the collaborative atmosphere at the event helps build an interdisciplinary community among researchers interested in making their own research reproducible.

The first ReproHack in the Netherlands took place on November 30th, 2019, co-organised by the CSD at the LU Library (see 'website[6]' and 'blog[7]'). In this paper, we will reflect on the organisation and output of the ReproHack format, and the role libraries can play in the process. Finally, we will provide recommendations for reproducibility practices based on participants' feedback and provide resources by which support centers such as Libraries can organize their own ReproHack together with researchers.

Event setting

The event was hosted at the LU Library and library staff was actively involved in the organisation. By providing a location unrelated to a specific department or research field, the organizers hoped to attract a diverse group of attendants. The organizing team itself had diverse backgrounds and consisted of two researchers from University Medical Centers, one researcher at a Computer Science Institute, a student at the Social Science Faculty and a library staff member. Information about the event was spread via a twitter account, newsletters, public agendas and the organizers' individual channels, such as at field-specific meetings, workshops or by word of mouth. As the meeting rooms were offered free of charge, the organizers had spare funds to provide lunch and drinks for the participants and print stickers as well as posters. The full day event was held on a Saturday. Forty-four participants attended the event and brought in their experiences from the fields of psychology, engineering, biomedicine, and computer science.

Event structure

Two speakers framed the event, the first one introducing participants to current developments on tools for reproducible research and the second one putting reproducibility into the broader context of open science. In between, participants worked on 'hacking' scientific papers by testing and troubleshooting the papers' reproducibility and documenting what hurdles they encountered. The event was spontaneously extended with an improvised hands-on tutorial on how to make computational code citable. During finalizing drinks and bites, participants further discussed their experiences of the day and got more opportunity to build their network. The schedule of the day can be found in Box 1.

10:00 Coffee and Tea

10:30 Welcome

10:40 'Reproducibility: The value in the practice' by Dr. Anna Krystalli

11:30 Forming groups and start hackathon

12:30 Lunch buffet

14:30 'A vision for Open Science beyond the reproducibility crisis' by Dr. John Boy

15:00 Continue hacking

16:30 Drinks and bites

Box 1. Schedule of the ReprohackNL day.

Feedback on the papers

Attendees reported their findings on the attempt of reproducing the chosen papers using a standardized feedback form (Table 1). Forms were returned by email to all authors.

	Question	*Answer type*
1	Name of participant	Text field
2	Which paper did you attempt to reproduce?	Text field
3	Did you manage to reproduce it?	Yes/No/Almost

4	On a scale from 1 to 10, how much of the paper did you manage to reproduce?	Scale 1 (none of it) to 10 (all of it)
5	Briefly, describe the procedure followed/tools used to reproduce it	Text field
6	What were the positive features of this approach?	Text field
7	Any other comments/suggestions on the reproducibility approach?	Text field
8	How well was the material documented?	Scale 1 (Hard to navigate) to 10 (Very well)
9	How could the documentation be improved?	Text field
10	What did you like about the documentation?	Text field
11	After attempting to reproduce, how familiar do you feel with code and method used in the paper?	Scale 1 (Not at all familiar) to 10 (Fully walked through the analysis code)
12	Any suggestions on how the analysis could be more transparent?	Text field
13	Rate the project on the reusability of the material	Scale 1 (not reusable) to 10 (easily reusable)
14	Are material clearly covered by a permissive enough license to build on?	Checkbox Permissive license for DATA included Checkbox Permissive license for CODE included
15	Any suggestions on how the project could be more reusable?	Text field
16	Any final comments?	Text field
17	Contact email	Text field

Table 1. Feedback form to be filled in by participants.

Thirty-one papers on different topics were available for reproducibility checks during the event. The papers were not curated by the organisation. The only inclusion criteria for papers were that authors agreed to their use during the event, and that associated code and data were available online. Out of the 31 available papers, 19 papers were freely chosen by participants to work on during the event and 24 feedback forms were provided to their authors. Five of the 19 papers were reported as successfully

reproduced (answer 'yes' to question 3) and six were reported as almost reproduced (answer 'almost' to question 3). Feedback form scores are shown in Figure 1.

The organisation made an effort to frame the event as a learning opportunity rather than a performance assessment, hoping to make this an accessible event also for researchers with less experience with reproducible workflows. Despite this communication strategy, it remained a major challenge to convince authors to submit their work for the event. The top three reasons why contacted researchers were hesitant to share their work for reproducibility checks were: 'My code is not nice enough to be shared', 'I work with sensitive data that cannot be shared', 'I don't want to think about this work I did two years back anymore'. These responses speak to the organisers' impression that, unfortunately, openly sharing code and data remains an uncomfortable thing to do within the current academic climate.

Issues related to reproducibility

Reproducibility (The Turing Way Community, et al., 2019) is the extent to which analysis results can be reproduced given the code and the exact same data as in the original analysis.

Reproducibility scores ranged from 1 (not reproducible) to 10 (fully reproducible), and 12 feedback forms reported a score higher than 5 (question 4). When participants were asked to rate the documentation and annotation of code, 21 feedback forms reported a score higher than 5 (question 10). Thus, there were more documentation scores higher than 5, than reproducibility scores higher than 5. In agreement with these findings, participants reported a number of issues other than documentation in the openly phrased feedback questions (question 5, question 7 and question 12), which caused them to be unable to reproduce the paper: unavailability of data (locked behind a paywall, not available due to privacy issues, time-outs during download), faulty code, and dependence of analyses on proprietary or platform-specific software. Various recommendations for improving code documentation (question 9) were provided. First, it was suggested to include a codebook, as often used in survey research, which provides information about the data structure by explaining the survey structure and aids data interpretation and transparency. A second recommendation was adding a README text file, which explains the nuances and context of a unique data collection and has been deposited to a data repository. Further, suggestions included better explanation of how the different parts of the code correspond to results in the paper, more and clearer comments in the code itself, performing a check for typos, and providing orientation on time needed for the code to run.

Issues related to reusability

Reusability (The Turing Way Community, et al., 2019) is the extent to which the code and provided data can be used in future studies and how easily the material can be built upon by other researchers.

The reported scores for reusability varied between 1 (not reusable) and 10 (easily reusable), and 15 feedback forms reported a score higher than 5 (question 13). Less than half (10) of the feedback forms reported the existence of a permissive license for either data or code (question 14). When participants were asked how familiar they were with the code and method after trying to reproduce, answers varied between 1 and 10, with 12 reporting a score higher than 5 (question 11). When asked to provide

suggestions to further improve the reusability, adding a permissive license was the most frequent answer (6 times) (question 15).
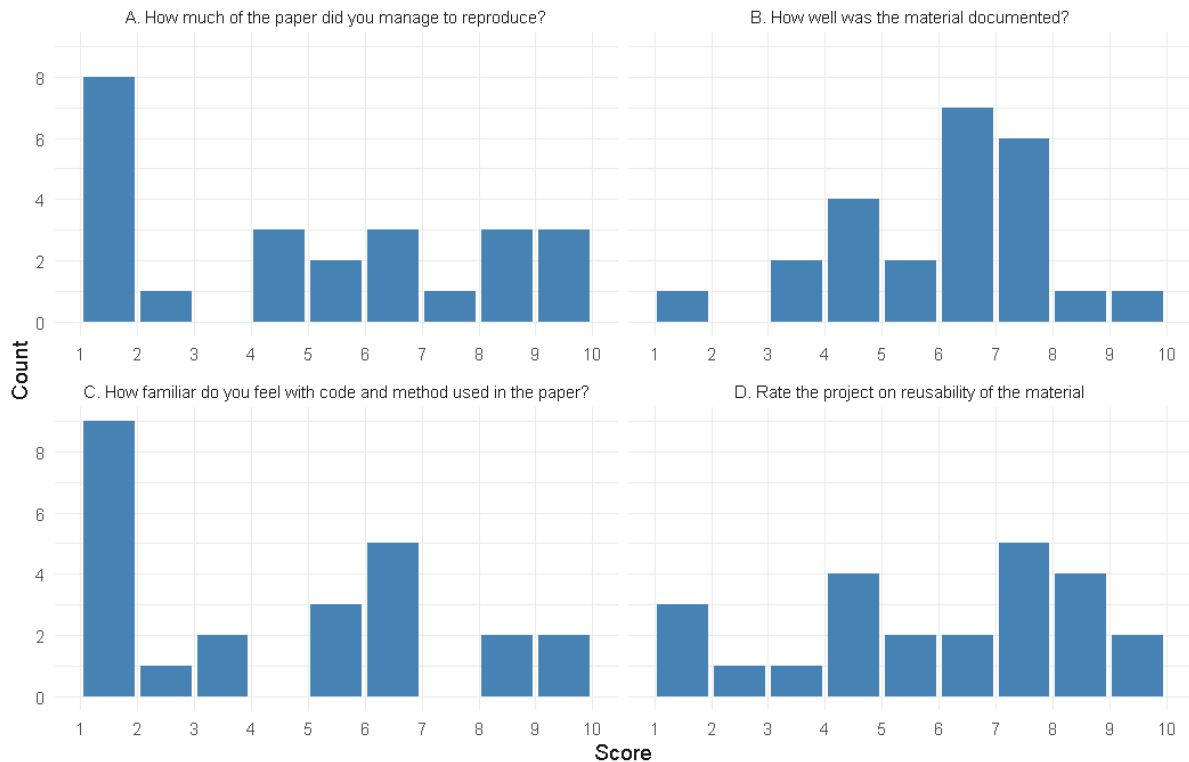

Figure 1. Distribution of feedback form scores for likert-type questions 4 (A), 8 (B), 11 (C) and 13 (D).

Comparison across teams

Three papers were reproduced by more than one team each, which allows for comparison of scores between teams. We compared the feedback form scores from the questions that could be quantatively compared (questions 4, 8, 11, and 13), see Table 2. We noted that scores for question 4 and question 13 varied the most (between 1 and 9), but also the answers to question 8 and question 11 varied, between 4 and 8, and between 1 and 7, respectively. The difference in scores for the third paper (Paper C in Table 2) might be related to the coding experience of the participants. From the text answers we could extract that one of the teams was not able to get the code to run at all and spend a lot of time on setting up the required software, while another team could solve R dependency issues more readily and work through the project.

Question	Answers from feedback forms:		
	Paper A	Paper B	Paper C
(4) On a scale from 1 to 10, how much of the paper did you manage to reproduce?	5/9/8	1/1	7/1
(8) How well was the material documented?	8/6/7	4/5	7/6
(11) After attempting to reproduce, how familiar do you feel with code and method used in the paper?	7/7/6	1/3	7/1
(13) Rate the project on the reusability of the material	9/8/8	4/5	7/1

Table 2. Feedback form scores per team for papers that were reproduced by more than one team. Read as score from team 1/ team 2/ team 3.

Language used in feedback forms

We noted that the language used in feedback forms was friendly and constructive, reflecting the interpersonal atmosphere at the event. To illustrate, Figure 2 shows a word analysis of question 15, containing more positive (blue words) than negative words (green words).

Figure 2. Word analysis of question 15, larger letters reflect higher frequency of the used positive (blue) and negative (green) words.

Discussion

At ReproHack, participants could learn about both good and bad practices for writing a reproducible paper and gain experience in using tools like 'Github[8]' for code development and 'Zenodo[9]' for depositing code. Making data and code available, using interoperable and open software, and documenting code well, are often advocated as the main drivers (Baker, 2016) for reproducibility. Participants could experience the nuances of these drivers, in that data availability is also about packaging the data so that it is easy and fast to download, and that code availability is also about using platform independent, freely available coding software, and attaching a codebook or contextual explanations about the data collection process. We believe that by participating in a ReproHack, researchers become more aware of potential barriers to reproducibility. In addition, we believe that the hands-on, community-driven format of ReproHack also helps to normalize making mistakes, and therefore normalizing peer-checking and collaboration. Lessons learned can be communicated to others but also help researchers in writing more reproducible papers themselves. Finally, participants are encouraged to reuse the ReproHack format and re-run the event to their home institutions.

Potential ReproHack organizers might want to discuss with each other which day of the week they want to pick for the event. In the Netherlands, a Saturday attracted young scholars from across the country. Choosing a Saturday instead of a weekday might also be easier to organize as meeting rooms are free (if buildings are open) and participants do not have any work duties on that day. The last point is an issue that we are very much aware of. To make reproducibility and open science part of mainstream research, events like these need to be treated like any other training or workshop by supervisors. We suggest listening closely to what the local open science community prefers and to cater to their needs. If library training programs are a fixed part of graduate school curricula, a reprohack could become part of top-down training and hence be held during working hours. There is also a danger of only getting early career researchers already actively involved in Open Science communities. This challenge can be addressed by organizing smaller and regular reprohacks, for example within a research group, in addition to or replacing regular presentations of intermediate or final results. This will normalize reproducibility practices and create a bigger audience for larger and interdisciplinary events. Finally, we recommend picking a location that offers Wifi, power outlets, and space for participants to break out into smaller working groups or have lunch, and found the library especially suitable while not burdening the limited funding available to the organisers.

There are more community initiatives to promote reproducible research, such as the collaborative book project The Turing Way (The Turing Way Community, et al., 2019), the 'Cure-Tier workshop[10]' aimed at librarians, archivists and information professionals, and the 'CODECHECK[11]' workflow and guidelines for checking code behind research papers. ReproHack distinguishes itself from these activities by providing a live setting in which researchers with different levels of experience with computationally heavy research and reproducible analysis can learn together and from each other. The feedback forms provided at Reprohack guide participants on what to expect from papers in terms of reproducibility and reusability; participants can incorporate these ideas in their own work. In addition, feedback forms are a possitive reinforcement to authors who shared their work. A topic for future research is to identify factors contributing to how reproducibility is assessed and how much the reviewer's own coding experience impact the reproducibility score.

Conclusion

Based on the feedback forms from participants, a list of top 10 tips to enhance the reproducibility and reusability of a paper can be created (Box 2).

- Package data so that it is easy and fast to download

- Provide non-platform specific code that is written using an open software

- Use a codebook explaining the data structure

- Include a README text file to explain the context of data collection

- Comment the code generously

- Perform a typo check

- Report on time needed to run the code

- Explain which parts of the code corresponds to which results in the paper

- Attach a permissive license to code

- Attach a permissive license to data

Box 2. Top 10 tips from ReproHack participants to enhance the reproducibility and reusability of a paper.

Libraries have always been a space for people to meet and exchange ideas. In a sense, they offer a 'neutral' space outside research institutes for researchers to focus on sub-parts of work. ReproHacks and other grassroots initiatives need exactly that: a place to meet, work, think and discuss. Libraries are connected within the faculties and they can use their network to reach researchers throughout the university. For this first ReproHack in the Netherlands, the CDS contributed greatly by offering its infrastructure and enhancing the organisers' outreach through posters, flyers and its Twitter account, but also informing their faculty liaisons to give them an opportunity to spread the word via their channels as well. Next time, more directed advertisements can be made by informing participants taking part in other workshops organized by the CDS.

ReproHack sparked discussions at other Dutch universities around organizing their own ReproHack. This process is facilitated by information on how to run a ReproHack provided in a 'Github repository[5]' and there are plans to organize webinars for those interested in running their own reprohack.

References

Baker, M., 2016. 1,500 scientists lift the lid on reproducibility. *Nature,* 533(7604), p. 452–454.

Jacobsen, A. et al., 2019. FAIR Principles: Interpretations and Implementation Considerations. *Data Intelligence,* 2(1-2).

Pashler, H. & Wagenmakers, E., 2012. Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence?. *Perspectives on Psychological Science,* 7(6), p. 528–530.

The Turing Way Community, et al., 2019. *The Turing Way: A Handbook for Reproducible Data Science.* https://doi.org/10.5281/zenodo.3233853

Wilkinson, M. D. et al., 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data,* 3(1), p. 160018.

---

[1] Contact author: Kristina M. Hettne; Leiden University Libraries, Centre for Digital Scholarschip; Witte Singel 27, 2311 BG, Leiden, The Netherlands; k.m.hettne@library.leidenuniv.nl. Affiliation and address for authors: Ricarda Proppert; Leiden University, Faculty of Social and Behavioural Sciences, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands; Linda Nab; Leiden University Medical Center, Department of Clinical Epidemiology, Albinusdreef 2, 2333 ZA Leiden, the Netherlands; L. Paloma Rojas-Saunero; Erasmus MC, Epidemiology Department, Doctor Molewaterplein 40, 3015 GD, Rotterdam, The Netherlands; Daniela Gawehns; Leiden University, Leiden Institute of Advanced Computer Science, Niels Bohrweg 1, 2333 CA Leiden, The Netherlands.

[2] https://www.library.universiteitleiden.nl/research-and-publishing/centre-for-digital-scholarship/workshops

[3] https://digitalscholarshipleiden.nl/articles/durable-access-to-book-historical-data

[4] https://github.com/DorienHuijser/DecisionTreeMRIData

[5] https://github.com/reprohack/reprohack-hq

[6] https://reprohacknl.github.io/ReproHack/

[7] https://www.software.ac.uk/blog/2020-01-15-reproducibility-hackathon-netherlands-aftermath

[8] https://github.com

[9] https://zenodo.org

[10] https://www.projecttier.org/fellowships-and-workshops/cure-tier-workshop/#about-the-cure-tier-workshop

[11] https://codecheck.org.uk