AGILE: GIScience Series, 5, 43, 2024. https://doi.org/10.5194/agile-giss-5-43-2024
Proceedings of the 27th AGILE Conference on Geographic Information Science, 4–7 Sept. 2024.
Editors: Alison Heppenstall, Mingshu Wang, Urska Demsar, Rob Lemmens, and Jing Yao.
This contribution underwent peer review based on a full paper submission.

Educating for Open and Reproducible Research in the Geosciences: Lessons from an MSc program

Frank O. Ostermann [1]

[1] Department of Geoinformation Processing, Faculty of Geographic Information Science and Earth Observation (ITC), University of Twente, The Netherlands

Correspondence: Frank O. Ostermann (f.o.ostermann@utwente.nl)

Abstract. Reproducibility and replicability of scientific research are challenges for today's data-driven (geospatial) research. While several initiatives aim to improve published research's reproducibility, comparatively little attention has been paid to ensuring that tomorrow's researchers learn about the importance of reproducibility and develop the skillset and practice to include it from the beginning in their work. To address this shortcoming, this paper presents a recent case study of a course at the MSc level, where students were given a range of published research studies to choose from with the aim to reproduce or replicate them, following best practices such as preregistration. The outcomes of the course are encouraging and suggest a wider adoption of such approaches in higher education. This paper aims to provide all relevant information to enable lecturers to experiment with this type of course.

Keywords. Reproducibility, higher education, challenge-based learning, scientific literacy

1 Learning reproducible research practices

Research has shown that also the Geosciences have a reproducibility crisis: The reproducibility of published (conference) papers is low (Ostermann et al. 2021; Kedron et al. 2019). This is frequently due to licensing issues with the input data or incompatibilities between computational environments. Journals as well as conference series have taken steps to address these issues, e.g., the AGILE reproducibility guidelines and reproducibility review (inspired by the CodeCheck initiative, https://codecheck.org.uk/).

Another important reason mentioned by authors is a lack of skills and knowledge (Nüst et al. 2018). A lack of ingrained practice also could play an important role. These two sources of irreproducible research suggest having a look at the training of future researchers at institutes of higher education.

All MSc programs have some form of academic skills training in their curricula since MSc graduates are the pool from which future researchers develop. For many years, the content of such academic skills courses was traditional and conservative: avoid plagiarism, handle citations well, etc. In the past decade, the opportunities and the push for open science, as well as the increase in published quantitative research, have made it difficult for course designers and lecturers to keep up and renew their academic skills courses.

From the concrete needs of a particular institution and curriculum, the idea was hatched to add a data mastery course to the curriculum (see next section for background). The main objective was that students gain more skills in handling geospatial data programmatically, i.e., with code.

From earlier work on teaching reproducibility (Ostermann 2021a), the idea was further developed to combine that objective with teaching the students about open science and reproducibility while making the curriculum's didactical approach more challenge-based. The following sections describe the design considerations, the concrete course set-up, and the lessons learned.

2 Setting-up and running the course

2.1 The context: curriculum and students

The target MSc program of Spatial Engineering (https://www.utwente.nl/en/education/master/programmes/spatial-engineering/) aims to provide students with a broad range of skills in the three knowledge domains of spatial information science, governance and planning, and technical engineering.

A key distinctive feature of the program are larger courses of 15 EC that span 10 weeks, during which students work in groups on a wicked problem (such as predicting, mitigating, and managing outcomes of human-induced earthquakes), reach out to actual stakeholders to integrate their views, knowledge, and needs, and use a wide range of skills, methods, and tools.

The students enrolling in this program come from a wide range of backgrounds: geographically from Europe, Africa, Asia, and the Americas, but also from varied domains, e.g., geoinformatics, civil engineering, liberal arts, or planning and design.

The past years have shown that Spatial Engineering students are very motivated and strong self-directed learners. However, student evaluations from past years also have shown that many students would appreciate more robust and extensive content about handling and analysing geospatial data, especially programmatically.

Thus, the decision was made to offer a mandatory course on Data Mastery for a total of 10 EC: Students would join an existing 7 EC companion course that serves as introduction to programming, covering sub-themes such as algorithmic/computational thinking (how to translate a problem into pseudocode), Python programming, and SQL basics (for storing geospatial data in a PostGIS database). Complementing that course is a newly designed 3 EC course for using their new skills on a real-world research challenge – the Data Mastery Challenge course covered in this paper. A key design decision was to focus the learning objectives on reproducibility and replicability, for the reasons outlined in the introduction.

2.2 Design considerations for the course

The design had to address two main questions: First, how to design a challenge-based learning course within the curriculum's context, and second, how to teach reproducibility to a diverse student population.

The author and course designer had several years of experience as coordinator of one of the 15 EC challenge courses, so the first question was tackled by building mostly on existing experience. The answer to the second question could build on prior work of designing tutorials and materials for an introduction to reproducibility. Thus, the main issue was how to scale up that second part and make it more challenge-based.

The first step was, unsurprisingly, a general internet search for insights from other higher education courses teaching reproducibility, as well as Google Scholar and Scopus search for recent literature.

The query used was: "reproducibility AND teaching AND higher education". Because the aim was not a systematic literature review, not every potential result was investigated thoroughly.

However, there were surprisingly few candidate studies, partially due to a lack of a search engine or registry dedicated to open educational resources. Some interesting findings include:

On a more general level, Pownall et al. (2023) present strong evidence from the literature that embedding reproducible research into education in a "hands-on" way has the potential to increase students' scientific literacy, engagement, and future attitudes. From their own teaching experience, Ostblom and Timbers (2022) identify three key strategies for successfully teaching reproducibility: (1) the motivation requires emphasis – why is reproducibility important and why should students want to learn and practice it; (2) guided instructions to help them get started and support them in deciding what to do (and what not); (3) lots and lots of practice. However, their approach is not challenge-based.

Ball (2023) presents a very structured limited exercise for undergraduate students, while Karathanasis et al. (2022) describe an actual course where students reproduce research work. However, it only required them to reproduce specific figures from papers, working in small groups, with the total amount of time unclear.

Other relevant resources for developing the Data Mastery Challenge course were the materials developed and collected by the Framework for Open and Reproducible Research Training (FORRT, https://forrt.org/), the R for Geographic Data Science open educational materials (https://github.com/sdesabbata/r-for-geographic-data-science), the University of Zürich's Center for Reproducible Science (https://www.crs.uzh.ch/en.html), and the University of Washington's Reproducible Data Science course (https://canvas.uw.edu/courses/1354201/assignments/syllabus). This latter is close to what this case study was attempting, and its course coordinator was so kind to share his experiences in personal communication. A key take-away message was that overall, the Reproducible Data Science course worked well, but the students had to spend considerable time and effort on finding suitable challenges, and these also differed in difficulty, making the assessment a challenge of its own.

The two main conclusions of the review were: First, it is a feasible approach to use existing, published research as basis for the challenges. Second, challenge-based learning takes more time, and having the students look for challenges themselves will not be feasible within the 10-week time frame of our academic calendar.

2.3 Structure and learning objectives of the course

The main design decisions were to a) have introductory classes at the start of the course, spaced out over a longer time at lower density; b) provide students with a selection of candidate research studies that should be reproducible or replicable with their skills set and available time, and during the middle part of the course have them choose one, and from it develop their reproducibility or replicability challenge; c) dedicate the last two weeks, after the students had finished the 7 EC companion course, to tackling the challenge.

In the context of this course, reproduction means the repetition of an existing study: Same data, same methods, and therefore hopefully same (or similar enough) outcomes. The objective of a reproduction is to check whether the reported methods and results work, and to gain a better understanding for improving on it.

Replication means to check whether the findings of a study hold in different contexts, e.g., whether an observed effect is detectable at different scales or different locations (different data). Such a replication by the students could be a benefit for the authors of the original study by extending it and was the encouraged type of challenge.

In order to level the playing field as much as possible, the candidate studies had to meet the following criteria:

- A quantitative spatial analysis that can be automated (scripted) with Python to practice the skills learned in the companion course
- Either no published, ready-to-use Python code or no published, fully pre-processed data; publicly available but unprocessed (original) data is acceptable; any combination of searching, finding, and pre-processing should be part of the challenge. The aim was not to show the gold standard for reproducible research, but to provide a challenge to solve.
- More than one data set required (integration of different data sources is a learning objective)
- No deep domain knowledge required
- Published/reported/documented study (but not necessarily a journal paper)

We approached colleagues from the faculty for candidate studies to be offered as challenges, because this would also allow us to have them to be available as challenge owners for a supervision session of 60 minutes.

We settled on a total of 4 studies to offer: one study working mostly with remote sensing imagery on biodiversity (Khamila et al. 2023), one working with remote sensing imagery in an urban context (Aguilar and Kuffer 2020), one using social media and census data in an urban context (Ostermann 2021b), and another one using point data and digital elevation models (unpublished).

The three learning objectives were:

LO1 Develop a conceptual analysis workflow and data management and sharing plan

LO2 Share a reproducible package of the implemented workflow, containing all required data and code and sufficient documentation

LO3 Reflect on your approach and implementation, and evaluate the degree of success of your reproduction or replication

The students had to show their learning progress with three main deliverables:

First, to introduce them to best research practices, they had to submit a pre-registration using the Open Science Foundation’s template (Bowman et al. 2020), which had been modified for a more streamlined experience, and was not used for the final grade (but it would feed into the next deliverable). This preregistration was then discussed in class and peer feedback was provided.

Second, a report that contained:

- A conceptual workflow showing all data collection and (pre)-processing steps leading to clearly defined goals (LO1)
- A data management and sharing plan (LO1)
- An evaluation of the implementation and outcomes with respect to the conceptual workflow and goals (LO3)
- An individual reflection of each group member on their contributions and lessons learned. These reflections can disagree with each other. (LO3)

Third, a reproducible code and data package that contained:

- All data (size and licenses permitting), or detailed instructions, preferably executable, on how to obtain missing data. (LO2)
- All code as ready-to-run as possible (LO2)

For the course structure and content, see Table 1.

2.4 Data and Software Availability Section

As noted above, there is no repository or search engine for open educational resources (OER) that is comparable to our search engines for journal articles. One reason for that is that there is no established format, rules, or process for publishing course materials.

Further, the host university does not have a clear policy or suggested path concerning OER. All learning materials are copyrighted by the institution, not the instructor. Thus, the course materials are not open published but available upon request. All research studies used for the challenges are openly available.

The course outcomes (student projects) are graded and not available (although individual students have decided to publish their work, e.g., see https://www.linkedin.com/pulse/urban-space-mapping-challenge-azza-sawungrana-7ymrc).

Table 1: Course structure and topics; student input in **bold**

Course week and session	Content
1 - 1	Course Intro: Data Mastery as prerequisite for open and reproducible research; **skill survey**
2 - 2	Open science, reproducibility and replicability; assessing a published study; introduction of research studies to choose from
Break	*Holiday break and focus on companion course*
6 - 3	Recap open and reproducible research; preregistration; from work plan to workflow; tools
	Choice of research study, group formation, begin development of a challenge
7 - 4	Catch-up session, supervised collaborative work session on preregistration
	Hand-in preregistration
8 - 5	Interoperability, reproducibility tools (incl. Github);
	Peer-review of preregistrations
9 - 6	Question hour and tool support
9 - 7	**Mid-term group presentations**, peer feedback
10 - 8	Brief group reporting/update, peer feedback, tool support
10 - 9	**Final group presentations,** CodeCheck initiative**,** wrap up
	Hand-in of final report two days later

3 Lessons learned

3.1 Challenge accepted and mastered!

All groups successfully submitted preregistrations and reproducible packages of their work. All groups chose for some sort of replication of the original study, i.e., they changed data inputs and methods, but for different reasons:

One group chose a publicly available remote sensing imagery resource instead of the resource in the original paper, because the original source was not available anymore. Another group transferred the area of study to a different city to test for generalizability of the findings, which forced them also to change methods because of different input data semantics. A third group extended the original study and investigated a follow-up research suggestion.

All groups reflected on their outcomes with respect to the original study, and successfully assigned tasks and roles, depending on their interests and skills. All students reported in their individual reflections notable learning progress.

The grading of the course work included the reproduction of the students' data and code packages to test LO2. Two groups decided to share their work on Github, one group chose the electronic learning environment. The code was mostly in the form of well-documented Jupyter notebooks. Together with other documentation (read-me files, report), the assessment of the course work did not take more time than a traditional assessment (e.g., written or oral exam, or longer written report). All groups achieved at least a grading score of 85% (8.5 out of 10 in the Dutch grading system).

3.2 Issues mentioned by students

The companion course on programming was not able to completely level the playing field for the students: the differences in skills between them were still notable when they started to work on implementing their challenge.

Further, activities from other, parallel courses which prescribed a stricter timeline and more deliverables or assignments can easily interfere with a more openly structured course like this one.

3.3 Learning moments for the teacher

The students developed interesting solutions to some problems, e.g., the use of social group chat as means to share data quickly.

The reproduction of their work provided new insights into what can go wrong, including setting up a working Python virtual environment without dependency conflicts.

Finally, there was not enough time for a deeper reflection of what the outcomes meant for the scientific process, i.e., a systematic evaluation of the outcomes with respect to the original study.

3.4 Discussion and outlook for the future

The course outcomes support earlier findings about students' scientific literacy and engagement (Pownall et al. 2023) – but it remains to be seen whether there will be lasting impact on their practices! The recommendations from Ostblom and Timbers (2022) worked well for engaging the students.

The course was overall a success and will be repeated in the future. Some more emphasis will be placed on best practices of sharing a computational environment, e.g., environment and requirements, possibly extending into container (Docker) technology if time allows, or as bonus content for more advanced students.

Acknowledgements

The colleagues in the Reproducible Research @ AGILE initiative (https://reproducible-agile.github.io/) for encouragement and feedback.

Prof. Dr. Ben Marwick (University of Washington) for kindly sharing his experiences in running a similar course.

Dr. Stef de Sabbata (University of Leicester) for setting a gold standard example of open educational resources on reproducibility.

The MSc program management and challenge providers (Tiny Luiten, MSc; Dr. Rosa Aguilar; Dr. Thomas Groen) for making the course possible.

The Spatial Engineering students of class 2023/24 for their enthusiasm.

References

Aguilar, Rosa, and Monika Kuffer. 2020. 'Cloud Computation Using High-Resolution Images for Improving the SDG Indicator on Open Spaces'. *Remote Sensing* 12 (7): 1144. https://doi.org/10.3390/rs12071144.

Ball, Richard. 2023. '"Yes We Can!": A Practical Approach to Teaching Reproducibility to Undergraduates'. *Harvard Data Science Review* 5 (3). https://doi.org/10.1162/99608f92.9e002f7b.

Bowman, Sara D., Alexander Carl DeHaven, Timothy M. Errington, Tom Elis Hardwicke, David Thomas Mellor, Brian A. Nosek, and Courtney K. Soderberg. 2020. 'OSF Prereg Template'. Preprint. MetaArXiv. https://doi.org/10.31222/osf.io/epgjd.

Karathanasis, Nestoras, Daniel Hwang, Vibol Heng, Rimal Abhimannyu, Phillip Slogoff-Sevilla, Gina Buchel, Victoria Frisbie, Peiyao Li, Dafni Kryoneriti, and Isidore Rigoutsos. 2022. 'Reproducibility Efforts as a Teaching Tool: A Pilot Study'. *PLOS Computational Biology* 18 (11): e1010615. https://doi.org/10.1371/journal.pcbi.1010615.

Kedron, Peter, Amy E. Frazier, Andrew B. Trgovac, Trisalyn Nelson, and A. Stewart Fotheringham. 2019. 'Reproducibility and Replicability in Geographical Analysis'. *Geographical Analysis*, August, gean.12221. https://doi.org/10.1111/gean.12221.

Khamila, C. N., T. A. Groen, A. G. Toxopeus, L. Santini, M. Neumann, C. Van Swaay, and H. Sierdsema. 2023. 'There Is a Trade-off between Forest Productivity and Animal Biodiversity in Europe'. *Biodiversity and Conservation* 32 (6): 1879–99. https://doi.org/10.1007/s10531-023-02582-2.

Nüst, D., C. Granell, B. Hofer, M. Konkol, F.O. Ostermann, R. Sileryte, and V. Cerutti. 2018. 'Reproducible Research and GIScience: An Evaluation Using AGILE Conference Papers'. *PeerJ* 2018 (7). https://doi.org/10.7717/peerj.5072.

Ostblom, Joel, and Tiffany Timbers. 2022. 'Opinionated Practices for Teaching Reproducibility: Motivation, Guided Instruction and Practice'. *Journal of Statistics and Data Science Education* 30 (3): 241–50. https://doi.org/10.1080/26939169.2022.2074922.

Ostermann, Frank O. 2021a. 'PEER ASSESSMENT TO IMPROVE REPRODUCIBILITY OF COMPUTATIONAL PROJECT WORK'. In , 1080–90.

———. 2021b. 'Linking Geosocial Sensing with the Socio-Demographic Fabric of Smart Cities'. *ISPRS International Journal of Geo-Information* 10 (2): 52. https://doi.org/10.3390/ijgi10020052.

Ostermann, Frank O., D. Nüst, C. Granell, B. Hofer, and M. Konkol. 2021. 'Reproducible Research and GIScience: An Evaluation Using GIScience Conference Papers'. In *Leibniz International Proceedings in Informatics, LIPIcs*, 208:VII. https://doi.org/10.4230/LIPIcs.GIScience.2021.II.2.

Pownall, Madeleine, Flavio Azevedo, Laura König, Hannah Slack, Thomas Evans, Zoe Flack, Sandra Grinschgl, et al. 2023. 'Teaching Open and Reproducible Scholarship: A Critical Review of the Evidence Base for Current Pedagogical Methods and Their Outcomes'. *Royal Society Open Science*, April. https://gala.gre.ac.uk/id/eprint/42425/.