OPEN ACCESS

Publiziert im Juni 2023
DOI: 10.15460/apimagazin.2023.4.2.144

Zeitschriftenartikel

Begutachtet

Begutachtet:

Dr. Steffen Rudolph
HAW Hamburg
Deutschland

Erhalten: *21. April 2023*

Akzeptiert: *25. Mai 2023*

Publiziert: *29. Juni 2023*

Empfohlene Zitierung:

FRICK, Claudia, BLÜMM, Mirjam, RANDALL, Natasha, KÜÇÜK, Berrak and BAILEY, Drew H., 2023: A reality check on research reproducibility in Open Science student's projects. In: *API Magazin* **4**(2) [Online] Verfügbar unter: DOI 10.15460/apimagazin.2023.4.2.144

A reality check on research reproducibility in Open Science student's projects

Claudia Frick[1]*, Mirjam Blümm[1], Natasha Randall[2], Berrak Küçük[3] und Drew H. Bailey[4]

[1] Technische Hochschule Köln, Deutschland
Professorin für Informationsdienstleistungen und Wissenschaftskommunikation

[2] Technische Hochschule Köln, Deutschland
Student*in im 2. Semester des Studiengangs Digital Sciences: Data and Information Science (MA)

[3] Technische Hochschule Köln, Deutschland
Student*in im 2. Semester des Studiengangs Digital Sciences: Software Architecture (MA)

[4] University of California, Irvine
Associate Professor

* Korrespondenz: redaktion-api@haw-hamburg.de

Abstract

Open Science as the foundation of transparent and reproducible science is increasingly being incorporated into curricula. We argue that Open Science education is predestined not to be taught in classical lectures, but to be experienced first-hand as reproduction studies in student projects. The case study of a successful Master's module is presented from three perspectives: the lecturers', the students' and the researchers' whose published study was reproduced. This illustrates that attempting to reproduce a published study is a very vivid and sustainable learning experience that naturally incorporates many Open Science topics, and that the students' work contributes to increasing the number of reproduced studies and ensuring the quality of the published body of scientific knowledge.

Keywords: Open Science, Reproducibility, Research Studies, Student's Projects, Research Data Management, Teaching

 DOI: 10.15460/apimagazin.2023.4.2.144

A reality check on research reproducibility in Open Science student's projects

Zusammenfassung

Open Science als Grundlage transparenter und reproduzierbarer Wissenschaft wird zunehmend in Lehrpläne aufgenommen. Unserer Auffassung nach ist Open Science prädestiniert dafür, nicht in klassischen Vorlesungen gelehrt, sondern als Reproduktionsstudien in studentischen Projekten aus erster Hand erlebt zu werden. Dafür stellen wir die Fallstudie eines erfolgreichen Mastermoduls aus drei Perspektiven dar: die der Dozierenden, die der Studierenden und die der Forschenden, deren veröffentlichte Studie reproduziert wurde. Es wird deutlich, dass der Versuch, eine veröffentlichte Studie zu reproduzieren, eine sehr nachhaltige Lernerfahrung ist, die ganz selbstverständlich viele Open-Science-Themen vereint, und dass die Arbeit der Studierenden dazu beiträgt, die Anzahl reproduzierter Studien zu erhöhen sowie die Qualität des veröffentlichten wissenschaftlichen Kenntnisstands zu sichern.

Schlagwörter: Open Science, Reproduzierbarkeit, Forschungsstudien, Studentische Projekte, Forschungsdatenmanagement, Lehre

 DOI: 10.15460/apimagazin.2023.4.2.144

1 Motivation

Open Science is the foundation of transparent and reproducible science.[1] Scientific studies should be described and conducted in a way that enables others to understand and reproduce them. It is not easy to judge at a first, second or even a third glance whether a study has been designed and described in such a way. Simply finding the associated data and code is not necessarily enough to make a reproduction possible. The next challenge is to design and describe one's own studies in a way that makes them reproducible.[2] Both perspectives, i.e. reproducing a study and creating a reproducibly study, are important. So far, reproducibility education tends to focus on researchers and not students (McAleer et al. 2022). The module "Open Science" in the Master's programme "Digital Science" at the Cologne University of Applied Sciences challenges that and considers both perspectives.[3]

The module is a combination of input and discussions centred around student projects. From the different concepts of reproducibility and replicability, to research data management and the development of research software, to the selection of licences and the communication of one's own results, everything along the way is covered. For this purpose, students reproduce self-selected studies and navigate their way through all the challenges, to either success or failure. Regardless of the outcome, this is a valuable and rewarding process, for students as well as the lecturers and the scientific community, and offers a long-lasting learning experience. In this article, we introduce the concept of the "Open Science" module and one of the students' projects from winter semester 2022/2023. The chapter on the teaching concept is written by the two lecturers (Claudia Frick and Mirjam Blümm) and the chapter on the students' project is written by the two students who conducted the reproducibility study (Natasha Randall and Berrak Küçük) amended by a statement of one of the authors of the original study (Drew Bailey).

2 Open Science teaching concept

The learning objectives of the "Open Science" module include the ability to lead a scientific discourse about Open Science, apply related tools and services, process and provide research data as well as understand and reproduce case studies. In terms of Bloom's taxonomy, students should achieve all cognitive domain levels as far as possible (Bloom et al. 1956). The course followed the "flipped classroom" principle (Kirch 2016) and was based on three main components: methodological

1 We define a study as reproducible if its results can be recreated with the same method and same data as used and published by the authors (Chiarelli et al. 2021, p. 10-11). A study is replicable if its results can be confirmed using the same method but new data, and is robust if its results can be confirmed using a new method but same data (Eickhoff 2020, 20:11-23:58). We acknowledge the many other possible definitions and discussions (Plesser 2018, van de Sandt 2019).

2 Even literature search can be described reproducibly (Booth et al. 2016).

3 See for further information https://digital-sciences.de/en/modules/open-science/ [Online, Accessed 21 March 2023].

DOI: 10.15460/apimagazin.2023.4.2.144

and technical content, classroom interaction, and practical student projects. In practice, the students prepared materials provided in advance via the learning platform Ilias[4] and in the first half (90 minutes) of the on-site lessons, essential content was picked up and jointly developed with the help of an interactive whiteboard. The second half of the lessons were reserved for group work on the student projects.

Fig. 1 shows an example unit on research data management (RDM) on the interactive whiteboard. After watching a short video about the difficulties of data reuse, students discussed use cases and compiled arguments for RDM, which they recorded with sticky notes on the whiteboard. We then discussed the different steps of RDM based on the research life cycle, and defined the overall tasks. As students had read the "Practical Guide to the International Alignment of Research Data Management" (Science Europe 2021) in preparation for the lesson, they already knew that a data management plan (DMP) is a key instrument for RDM. In class, the use and application of DMPs was reflected on, and its elements recorded on the whiteboard. In the subsequent group work, the students considered the importance of RDM for their data, and how they could use its elements for their projects.

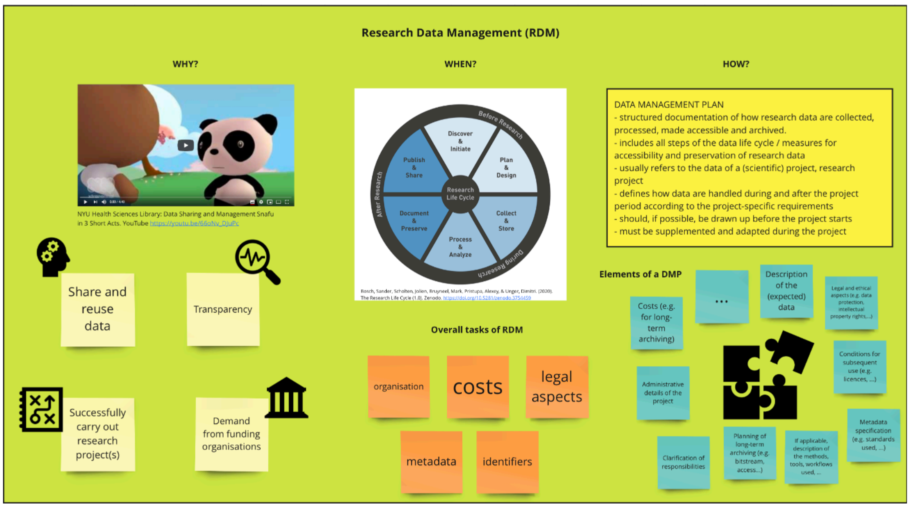

Fig. 1: Screenshot of the unit on Research Data Management on the interactive whiteboard.

In this spirit, we designed nine units for the module; the first unit contained a general introduction to the topics covered, and the group work started with the search for possible studies to reproduce. For this, research tools and methods were discussed in class. The second unit dealt with several aspects of RDM, including how data can be structured and organised. The group work continued after a discussion of approaches to search for and identify reproducible studies. Scientific publishing was the focus of the third unit. It dealt with the different methods of scholarly communication, the idea of open access, and the means by which research data can be published and retrieved. The student groups shortly summarised their selection criteria and results so far. The topic of RDM was continued in the fourth unit and

4 Available at https://www.ilias.de/ [Online, Accessed 21 March 2023].

DOI: 10.15460/apimagazin.2023.4.2.144

focused on the FAIR Principles (Wilkinson et al. 2016), data and metadata formats, as well as legal and ethical aspects. The student groups reported on their progress and continued to work on their reproducibility study and documentation.

The second Research Data Day in North Rhine Westphalia in 2022[5] was integrated into the schedule as the fifth unit. Students were encouraged to participate in the online programme, offering insight into various current FDM projects.[6] In the sixth unit each group presented an overview of the study they had chosen to reproduce, including the topic, research question, method, data, code, and result. They described their current progress (communication with the authors, coding or data usage, problems – no matter if they had solutions right now or not) and addressed their next steps. The seventh unit covered research software and infrastructure, and stressed the importance of open and sustainable software, as well as the difficulties arising from its maintenance. Afterwards, students continued with their projects, utilising the feedback on their presentations of the previous week. The eighth unit gave an overview of national and international players and initiatives in the field of Open Science. The group work concentrated on preparing the final presentation, which took place in the ninth and final unit.

For the student projects, three groups were formed and three different studies chosen (Ariyo et al. 2014, Xu et al. 2022, Li et al. 2014). One of the biggest challenges from our perspective was dealing with the different levels of uncertainty: will students identify original studies that are at least theoretically reproducible? Will the methodology and resources (e.g., data, code) be both available and understandable? Will students be able to reproduce the study? How will we handle communication if inconsistencies in the original studies arise? Despite various difficulties (such as missing metadata and outdated software), the students were able to reproduce their chosen studies to a large extent. A particularly good example, especially in terms of Open Science, is described in more detail in the next section.

3 Student project

Our first criterion when choosing a study for our project was to look for papers with publicly available datasets, as these were not so easy to find; many studies either do not publish their data, or the data format is not very accessible. Another criterion was that the study should be understandable even without specific domain knowledge. We conducted a search through DataCite[7] and discovered the study "Women's Preference for Masculine Traits Is Disrupted by Images of Male-on-Female

5 See https://web.archive.org/web/20221110095004/https://www.fdm.nrw/index.php/tag-der-forschungsdaten-in-nrw/tag-der-forschungsdaten-in-nrw-2022/ [Online, Accessed 21 March 2023].

6 See https://blog.rwth-aachen.de/forschungsdaten/2022/11/24/rueckblick-tdf-2022/ [Online, Accessed 21 March 2023].

7 See https://datacite.org [Online, Accessed 21 March 2023].

Aggression" (Li et al. 2014). The paper was published in PLOS One, and the raw data (Li et al. 2015) was stored on the open-access repository Dryad[8] and licensed under a CC0 1.0 licence, allowing for re-usability. We found the paper interesting, and it had been cited 26 times as of April 2023, according to PLOS One. We therefore decided to attempt to reproduce this study.

In the original experiment, 20 photographs of men's faces were graphically transformed into a pair of feminised and masculinised versions, and the 331 female participants chose which face they preferred. They were then shown their assigned group's priming images: of either male-on-female aggression (e.g., domestic violence), male-on-male aggression (e.g., boxing), male intergroup aggression (e.g., soldiers), neutral (e.g., reading a book), or pathogen (e.g., dirty toilet) images. The participants once again selected their preferred face. A linear random intercept model with a logistic response variable extracted the variance in the women's masculinity preferences, caused by the different priming groups. The study concluded that regardless of the priming images shown, participants tended to prefer the masculinised face, and this preference increased over time. The main finding of the study was a significant interaction (p=0.011) between time and the male-on-female aggression priming group; the preference for masculinity of the participants in the male-on-female aggression priming group had not increased over time. The study's authors had provided an email address with the paper, allowing us to contact them and state our intentions to attempt to reproduce their study. The immediate response was extremely positive, expressing interest in the outcome of our project, and a willingness to help in any way they could. However, at the time of our project the original paper had been published over 8 years ago, thus the authors struggled to recall many of the details of the study, hindered by a lack of documentation and metadata. Author Drew Bailey provided us with all of the available, additional files, including the original R code[9] used for the analysis. The code utilised now-obsolete functions and contained few comments; it was clear that the code had not been written for reuse by third parties, or for long-term maintenance. It quickly became evident to us, that even with a full dataset and original code available, reproduction of a study is very difficult without also having a comprehensive description of the study's methodologies.

We therefore strived to apply the open practices we had been taught in the Open Science course when carrying out our reproduction work. We created a data management plan, organised the files into structured directories, and created additional descriptive metadata. We considered FAIR principles (Wilkinson et al. 2016) when constructing a modified dataset, as well as appropriate, open licences for our data and research software. Having been made aware of the importance of open methodology (Besançon 2021), we made all of our code publicly available on

8 See https://datadryad.org/stash [Online, Accessed 21 March 2023].
9 See https://www.r-project.org/ [Online, Accessed 21 March 2023].

 DOI: 10.15460/apimagazin.2023.4.2.144

GitHub in heavily commented Jupyter notebooks, so that our reproduction work could be easily understood and in turn replicated by others.[10] Through our work, we demonstrated the reproducibility of the paper, as we were able to successfully recreate the study's main statistical model, and replicate the published results. We also constructed some of our own alternative models, to test for the robustness of the study; these also supported the conclusions of the original paper. Our results seemingly contributed to the body of successfully replicated studies, tackling the replication crisis in science (Baker 2016) – but reproducibility alone does not necessarily tell the whole story. When exploring the original dataset, we discovered that the data for 40 of the participants had accidentally been duplicated from other participants in the study. From reading a comment in the R code and following through its implications, we also realised that one of the face image pairs was miscoded, with the wrong side assigned as masculinised; hence the results relating to that particular image had been incorrectly inverted. Our next step was therefore to correct these errors – removing the 40 duplicated participants from the dataset, and inverting the wrongly assigned image data. We then recreated the study's main model with the fixed dataset. We were pleased to find that the conclusions of the paper remained valid, and the main finding was even strengthened as a result of the correction; the original p-value for the male-on-female priming condition interaction with time, of p=0.011, was now reduced to p=0.005. **Figure 2** depicts the original key figure in the paper (left), next to our equivalently scaled, corrected reproduction (right).

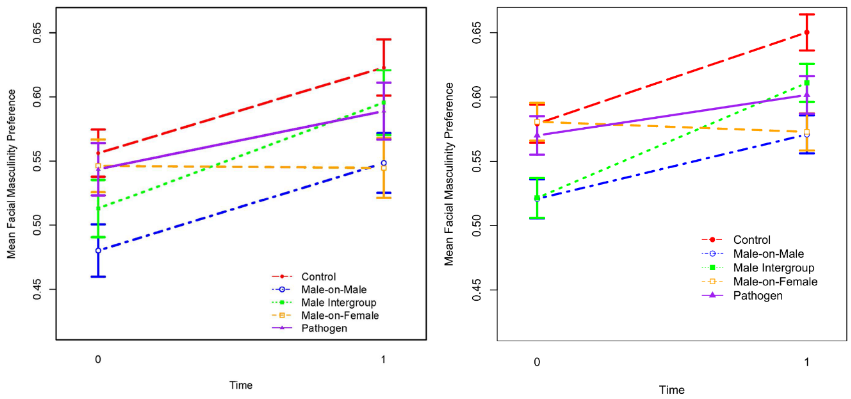

Fig. 2 Left: the original figure (Li et al. 2014). Right: our corrected reproduction.

Forms of open peer review have been suggested to induce insincerity through timidity in reviewers (Pros and cons of open peer review 1999), and after the paper's authors had been so generous in their willingness to help out with our project, we as students naturally had reservations about claiming to have found errors in their published paper. Nevertheless, we communicated the results of our reproduction back to the authors, and fortunately were met once again with an incredibly positive response. Drew Bailey encouraged us to write and submit a correction letter to the

10 See https://github.com/Natasha-R/Project-Reproduction-Open-Science [Online, Accessed 21 March 2023].

 DOI: 10.15460/apimagazin.2023.4.2.144

journal, referencing the findings of our project work. We also published the correction on PsyArXiv (Randall et al. 2023) and linked to it in a comment on the original study. This was endorsed in a responding comment by Drew Bailey, who continued to work with and support us throughout the process; his perspective is described in the following section.

4 Original author's perspective

When I received an email from Natasha and Berrak requesting some data and clarification from our 2014 study on masculinity preferences, I experienced mixed emotions. On the one hand, it is flattering to learn that students are interested enough in one's work to take the time to download the data and reproduce it. On the other hand, I admit the idea that an independent team is reexamining one's previously published work is also a little scary.

Yet, the only appropriate response for me was to offer them my fullest support. I opened the R code I had written for the project and became disappointed, for it contained little documentation at all. This was one of the first projects – I was a Ph.D. student and the reproducibility movement in psychology was unfolding between the time we designed the study (2011), analysed the data (sometime between 2012 and 2013), and published the paper (2014) – for which we had published our data. We took some pride in our transparency then, but I had no formal training in Open Science practices. In hindsight, it seems silly to think that anyone would ever publish raw data without clearly commented code, but I had not performed a reanalysis of data from a previously published paper at the time, and did not sufficiently consider what kind of information would necessitate such an undertaking.

I let my co-authors know right away about the students' project. Fortunately, they were very supportive. When the students found two clear errors – duplicated participants that should have been removed when data were pulled at the end of the study, and a miscoded item – I verified them in my data and code, reran the analysis, and informed my co-authors about the errors, taking full credit for them. I was, again, disappointed, but took some solace in finding the results were largely unchanged. I think a reason for this is that the results reported in the paper were not selected on statistical significance; indeed, we reported mostly null estimates in the paper – as I recall, it was rejected at another journal partially because of this. If we had, then errors would be correlated on average with statistical significance, and corrections would have been more likely to invalidate our findings. To be clear, this is not a defence of sloppy data management practices, but it is yet another reason that selecting which estimates to report prior to seeing them is good practice.

Finally, I felt responsible to make sure the correction was published. I encouraged Natasha and Berrak to reproduce the model results with the corrected dataset, and

DOI: 10.15460/apimagazin.2023.4.2.144

I reproduced their estimates with my own code. They wrote up a short, clear correction and posted it on a preprint server. They added a comment to the website on which the original article is posted, and I replied to it encouraging anyone interested in citing our paper to cite their correction as well.[11]

Despite intellectual, technical, and structural innovations designed to facilitate replication and reproducibility, these tasks remain time-consuming and difficult: even with my full cooperation, it took approximately five months since I received Natasha and Berrak's first email for them to post their correction online. Although I view structural factors as important for the past and future success of Open Science (e.g., Freese, Peterson 2018), social interactions have and will continue to play an important role in the proliferation of Open Science practices (e.g., Janz, Freese 2021). Scientists, particularly those who received most of their training prior to the Open Science movement, should take responsibility to minimise the still-present social barriers to replicability and reproducibility.

5 Discussion

Open Science education is predestined not to be taught in classical lectures and seminars, but to be experienced first-hand in student projects. Two reasons for this are well illustrated by our case study. Firstly, we live in a world where reproduction of published results is still all too rare. Therefore, student projects attempting to reproduce published studies contribute to closing this gap, which is motivating for the students as well as the lecturers. Moreover, it enables students to connect with published researchers, and challenges everyone involved to communicate about possible corrections. Secondly, trying to reproduce a published study is a very vivid and long-lasting learning experience, because one experiences first-hand what a study needs, to have even a chance to be reproduced in the first place. Additionally, published studies are self-contained units which naturally define individual student projects, and discussing their suitability at the beginning and during the module certainly brings to light a lot of Open Science topics. All in all, this approach can be carried out successfully, but it requires a lot of flexibility and commitment from teachers, students and the scientific community as a whole.

11 See https://journals.plos.org/plosone/article/comment?id=10.1371/annotation/920fbfb0-2d43-41f9-87a9-60784e79a95c [Online, Accessed 21 March 2023].

References

ARIYO, Adebiyi A., ADEWUMI, Adewumi O. and AYO, Charles K., 2014. Stock Price Prediction Using the ARIMA Model. *UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, Cambridge, UK, 2014, pp. 106-112, DOI: 10.1109/UKSim.2014.67.

BAKER, Monya, 2016. 1,500 scientists lift the lid on reproducibility. *Nature.* 25 May 2016. 533(7604), pp.452–454. Available from: DOI: 10.1038/533452a.

BESANÇON, Lonni, PEIFFER-SMADJA, Nathan, SEGALAS, Corentin, JIANG, Haiting, MASUZZO, Paola, SMOUT, Cooper, BILLY, Eric, DEFORET, Maxime and LEYRAT, Clémence, 2021. Open science saves lives: lessons from the COVID-19 pandemic. BMC Medical Research Methodology. December 2021. Vol. 21, no. 1, p. 117. DOI: 10.1186/s12874-021-01304-y.

BLOOM, Benjamin S.; Krathwohl, David R. and BERTRAM, Masia B., 1956. *Taxonomy of educational objectives: The classification of educational goals.* Handbook I: Cognitive domain. New York: David McKay Company.

BOOTH, Andrew, SUTTON, Anthea and PAPAIOANNOU, Diana, 2016. *Systematic approaches to a successful literature review*. Second edition. Los Angeles: Sage. ISBN 978-1-4739-1245-8.

CHIARELLI, Andrea, LOFFREDA, Lucia and JOHNSON, Rob, 2021. *The Art of Publishing Reproducible Research Outputs: Supporting emerging practices through cultural and technological innovation.* Online. Zenodo. [Accessed 21 March 2023]. Available from: DOI: 10.5281/zenodo.5521077.

EICKHOFF, Simon, 2020. *Forschung im Fokus: Widersprüche in der Wissenschaft.* Online. Haus der Universität. [Accessed 21 March 2023]. Available from: https://youtu.be/3g1lp4NZHwM.

FREESE, Jeremy and PETERSON, David, 2018. The Emergence of Statistical Objectivity: Changing Ideas of Epistemic Vice and Virtue in Science. *Sociological Theory*. September 2018. Vol. 36, no. 3, p. 289–313. DOI: 10.1177/0735275118794987.

JANZ, Nicole and FREESE, Jeremy, 2021. Replicate Others as You Would Like to Be Replicated Yourself. *PS: Political Science & Politics*. April 2021. Vol. 54, no. 2, p. 305–308. DOI: 10.1017/S1049096520000943.

KIRCH, Crystal, 2016. *Flipping With Kirch: The Ups and Downs from Inside My Flipped Classroom*. New Berlin, Wisconsin: The Bretzmann Group. ISBN 978-0692661901.

LI, Yaoran, BAILEY, Drew H., WINEGARD, Benjamin, PUTS, David A., WELLING, Lisa L. M. and GEARY, David C., 2014. Women's Preference for Masculine Traits Is Disrupted by Images of Male-on-Female Aggression. PLoS ONE 9(10): e110497. DOI: 10.1371/journal.pone.0110497.

LI, Yaoran, BAILEY, Drew H., WINEGARD, Benjamin, PUTS, David A., WELLING, Lisa L. M. and GEARY, David C., 2015. Data from: Women's preference for masculine traits is disrupted by images of male-on-female aggression, Dryad, Dataset, DOI: 10.5061/dryad.9bg43.

MCALEER, Phil, STACK, Niamh, WOODS, Heather, DEBRUINE, Lisa, PATERSON, Helena, NORDMANN, Emily, KUEPPER-TETZEL, Carolina E. and BARR, Dale J., 2022. *Embedding Data Skills in Research Methods Education: Preparing Students for Reproducible Research*. Online. 3 November 2022. PsyArXiv. [Accessed 21 March 2023]. Available from: DOI: 10.31234/osf.io/hq68s.

PLESSER, Hans E., 2018. Reproducibility vs. Replicability: A Brief History of a Confused Terminology. *Frontiers in Neuroinformatics*. 18 January 2018. Vol. 11, p. 76. Available from: DOI: 10.3389/fninf.2017.00076.

Pros and cons of open peer review, 1999. *Nature Neuroscience*. Vol. 2, no. 3, p. 197–198. DOI: 10.1038/6295.

RANDALL, Natasha, KÜÇÜK, Berrak, LI, Yaoran, BAILEY, Drew H, WINEGARD, Benjamin, PUTS, David A., WELLING, Lisa L. M. and GEARY, David C., 2023. *Correction to: Women's Preference for Masculine Traits Is Disrupted by Images of Male-on-Female Aggression*. Online. preprint. 30 March 2023. PsyArXiv. DOI: 10.31234/osf.io/ap38y.

SCIENCE EUROPE, 2021. *Practical Guide to the International Alignment of Research Data Management – Extended Edition.* 27 January 2021. Available from: DOI: 10.5281/zenodo.4915862.

VAN DE SANDT, Stephanie, DALLMEIER-TIESSEN, Sünje, LAVASA, Artemis and PETRAS, Vivien, 2019. The Definition of Reuse. *Data Science Journal*. 20 June 2019. Vol. 18, p. 22. Available from: DOI: 10.5334/dsj-2019-022.

WILKINSON, Mark D., DUMONTIER, Michel, AALBERSBERG, IJsbrand J., et al., 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data.* 15 March 2016. 3(1), pp.1-9. Available from: DOI: 10.1038/sdata.2016.18.

XU, Kevin and NGUYEN, Hieu, 2022. Predicting housing prices and analyzing real estate market in the Chicago suburbs using Machine Learning. DOI: 10.48550/arXiv.2210.0.