Perspective
Reproducibility Starts from You Today

Yasemin Turkylmaz-van der Velden, 1, * Nicolas Dintzner, 1 and Marta Teperek 1
1 Delft University of Technology, Mekelweg 5, 2628 CD Delft, the Netherlands
*Correspondence: y.turkyilmaz-vandervelden@tudelft.nl
https://doi.org/10.1016/j.patter.2020.100099

SUMMARY
Who hasn’t yet heard about the debates on research reproducibility, or, perhaps even more, about the research reproducibility crisis? There have been numerous papers in the past several years discussing reproducibility issues in research. In addition, funders, publishers, and research institutions followed policies aiming at increasing research reproducibility. But what does it mean in practice for research to be reproducible? And where does one start in this flood of information, tools, and requirements? In this article, we aim to help researchers improve the reproducibility of their work by providing simple tips and good practices that can be readily applied at different stages of the research life cycle. Reproducibility starts from you. Today!

INTRODUCTION
Ninety percent of 1,576 researchers who participated in Nature’s survey on research reproducibility agreed that there was a reproducibility crisis in science. 1 According to this survey, up to 85% of the participants have failed to reproduce someone else’s results, while up to 65% of the participants have failed to reproduce their own results. Respondents thought that intense competition, and practices such as “selective reporting” and “pressure to publish” were top contributors to irreproducible research. These were followed by factors such as “methods, code unavailable” and “raw data not available from original lab,” suggesting that sharing is not common practice in the current research culture.

Research reproducibility has been on the radar for some time 2-4 and is certainly gaining more attention, as can be seen in changes in funders 5-7 and publishers 6, 7 policies regarding data availability and the emergence of the FAIR data principles. 8
FAIR data principles require data to be Findable, Accessible, Interoperable, and Reusable. The need for making these principles a requirement has been facilitated by the outcomes of various recent studies showing that data initially claimed to be available upon request are actually not available. 9, 10 This is not surprising given that in the current publishing culture the common practice is to disseminate research results by only publishing representative figures, or figures which are basically snapshots of finalized datasets, such as plots. These figures do not reveal the original data that has enabled their generation in the first place, nor do they explain how the data were analyzed. Additionally they are often not suitable for reuse, validation, and reproduction.

In this article, we provide some simple tips to help researchers increase the reproducibility of their research practices regardless of their domain. For the purpose of this article, we chose to use the term “reproducibility” to describe scientific activities aiming to obtain identical or similar results of a previously executed experimental or computational research activity in an identical or similar context. This article follows the life cycle of a research project. We start by outlining reproducibility concerns that should be addressed early in a project, followed by useful tools and practices to improve documentation and version control, and subsequently address the archiving of research outputs. We conclude with a discussion about reproducibility in a broader context.
PLANNING FOR REPRODUCIBILITY SHOULD START AT THE VERY BEGINNING OF THE PROJECT

Research reproducibility should be considered at the start of a research project. Otherwise, trying to establish all the connections between data, code, and the details of the methodology behind the experiments at the time of publishing papers describing the results might seem daunting if they have not been systematically recorded from the beginning.

So what are the tips for a good start?

Carefully Design Your Research Project
Think about your research question and think about the methodology you will apply to answer the question. Think also about the data you will collect. How are you planning to collect these data, how will you analyze the data, and what kind of tools and infrastructure will you need?

Write Up and Publish the Design of Your Study as a “Registered Report”
What this means is not only that someone will peer-review your study design and provide you with useful feedback, but you will also get a guarantee that the outcomes of your research will be published in a peer-reviewed journal, irrespective of the outcomes of your research. In other words, publishing your study design as a Registered Report really helps to tackle publication bias: all outcomes, “positive” and “negative,” are equally important for science to move forward. Publishing registered reports is a fairly recent practice but is already present in various domains such as humanities, medical research, and social sciences.

Start a Data Management Plan for Your Project
You might have heard about funders’ requirements for data management plans (DMPs). If you get public funding for your research, a written DMP is likely to be one of the conditions of your grant. However, starting with a DMP is also an excellent practice by which to improve your research reproducibility. More and more research institutions now customize their guidance for writing a DMP to provide researchers with information about the most useful storage, backup, and version control solutions, as well as other tools and platforms available to them. Therefore, while writing a DMP, you will also make important decisions about where and how the data are stored during the project, and how to keep track of and version your data, which will all have important implications on the reproducibility of your research. In addition, when creating a DMP you will also obtain useful references to access ethics or data privacy resources, which might be essential for your project (e.g., do you need ethical approval? Do you need secure storage for confidential data?). Moreover, already at the planning stage, you can address strategies for long-term data preservation: what to keep, what not to keep, who will have access to the data, and how. How will you share your data after the project, and what kind of tools and platforms will you apply to answer the question. Think also about the method-ology you will apply to answer the question. Think also about the data you will collect. How are you planning to collect these data, how will you analyze the data, and what kind of tools and infrastructure will you need?

Plan for Computational Reproducibility
Most of today’s research includes some form of computation, often in the form of research software (from full-fledged applications to smaller analysis scripts). If software is an important component of your project, you should also consider how it will contribute to the overall reproducibility of your work. Common pitfalls in this area are missing dependencies (third-party software you use and forget to indicate in the documentation of your research software or analysis scripts), environment issues (operating system compatibility), and overly complex installation processes. The choice of software tools and platforms will have a pervasive effect on how you run your experiment and how you will share your results. It is, therefore, wise to consider those early in the planning phase.

There are a number of existing solutions to mitigate these issues and thereby make it easier for others to reproduce your work. Dependency managers, which are tools dedicated to software library dependency tracking and installation (MAVEN for Java or PIP for Python, among others), can facilitate the identification and/or download of libraries. Containers (such as Docker) can ease the deployment of your tool. Containers allow you to work in an empty workspace, such that the execution of your software is not affected by everything that you previously installed on your computer. Although this forces you to document all your dependencies, you benefit from automated installation and execution facilities. If you are working on smaller projects, you might want to start with a notebook such as Jupyter Notebook. The advantage of using notebooks is that you can do everything in a single online environment: execute your scripts, and easily add notes and additional information. Using an online environment also means that you don’t have to install anything—you start straight away. Most of these tools are open source, actively developed and supported, and commonly used in industrial settings. They all have extensive online documentation, with tutorials provided by the community.

Ask for Help
This might all seem like quite a lot to do and feel intimidating, especially for those who are only starting with the whole reproducibility agenda. So don’t hesitate to ask for help. Start with your peers. Perhaps they could provide you with useful tips on how to improve the reproducibility of your work, recommend some useful tools, or agree to review your code. Your university might also have dedicated resources to support you. More and more institutions employ dedicated disciplinary data stewards and data managers to advise researchers on data management practices, as well as research software engineers to provide professional programming support. In addition, academic libraries often have dedicated resources to help with data management and research reproducibility (advice, training, tools, and infrastructure). Even if community members are unable to help you directly, they will most likely connect you with someone who could.

DURING YOUR PROJECT: DOCUMENTATION AND VERSION CONTROL IS ESSENTIAL FOR REPRODUCIBILITY

Documentation and version tracking of your experimental or computational protocols and setups are key for reproducibility.
Your first setup may not be optimal, and you are likely to make adjustments to it. Keeping track of all these changes will save you a lot of time—you will always know what worked best and why.

Document Your Experimental Work
If you are doing experimental work, it is important to write down all the details necessary for a future reader, a future colleague, and, most importantly, your future self to understand how you have carried out a certain experiment. There are various ways to do this. Typically, institutions require researchers to write everything down in a lab notebook. This can be done using a paper notebook; however, this makes it quite difficult to report details of data mostly captured in a digital setting. Also, what if your paper notebook gets accidentally lost or destroyed? Alternatively, you can use an electronic lab notebook (ELN). ELNs make it easy to digitally document your work, link your data and samples, record the parameters of experiments, or note down the results of any measurements. Moreover, ELNs are searchable and offer version control to make all your changes traceable. In addition to using an ELN, you can also add documentation about your data as a README.txt file (a plain text file) to the folder where the corresponding datasets are saved. Documentation may include details on the methodology used, analytical and procedural information, and data-specific information (such as parameters and/or variables, column headings, and codes/symbols).

Document Your Computational Work
Understanding code is difficult, especially if you did not write it yourself. Documenting your code will help you and others understand how the code works, both for validation purposes and to build on the existing code base. To document your code, you can rely on the documentation standards of the programming language you are using. This can be done using, for example, “Docstrings” in R, Python, or Javadoc format in Java. Dedicated integrated software development (such as RStudio, Eclipse, or VS Code) often facilitate the writing process of comments and provide plugins to generate documentation in PDF or HTML format.

There are many resources available online for such tool-supported practices. Therefore, by using an integrated development environment (most are free), you can easily increase the reproducibility of your computation and facilitate long-term reuse.

Use Version Control
Version control systems (such as Git, Mercurial, or Subversion) help you keep track of all changes made but also who performed them. Such tools also allow easy tracking of the different versions of a given software program and facilitate the identification of specific valuable revisions (e.g., the version used in a specific publication), as illustrated in Figure 1. Widely used in industry, especially for source code management, these tools have great potential in a research context. You can use them for software source code as well as data. Version control will prove particularly useful if collaboration with multiple partners is expected. As version control tools have been used intensively for more than a decade, there are many online resources available, including “how to start” tutorials.

AFTER YOUR PROJECT: HOW TO SHARE YOUR WORK SO THAT OTHERS CAN REPRODUCE IT

Research takes a lot of time, money, and effort, and not every researcher has access to the same facilities. Therefore, sharing research outputs enables others with limited resources to analyze and reuse them for new purposes. Additionally, science can only progress if we can build on top of each other’s results, and validation and reproduction of existing results is the first step.

Our first recommendation when it comes to archiving of research outputs is to consider the FAIR principles. The FAIR principles stand for Findability, Accessibility, Interoperability, and Reusability of research data. However, being FAIR is not a binary state, and one can strive to make their data “FAIRer.” In this section, we highlight what we consider to be critical for both long-term archiving and the FAIRness of research outputs while offering the most tangible selfish benefits.

So, how to share your work?

Archive Your Outputs
Instead of sharing upon request you can archive your outputs in a repository, which is a place where things can be stored and shared. There are repositories for archiving datasets, protocols, software, and other outputs. You can find an overview of available repositories for data archiving in re3data.org or FAIRsharing. Such platforms index both domain dedicated repositories (such as Gene Expression Omnibus, NOMAD Repository, and PANGAEA) and generic repositories, suitable for all domains (such as Zenodo or figshare).

Increase Your Citations
Perhaps the most important aspect of publishing outputs in a dedicated repository is obtaining a unique persistent identifier such as a Digital Object Identifier (DOI). This ensures that the archived outputs are always available online and prevents “404 not found” messages. Additionally, many repositories nowadays provide citation statements for the archived outputs, similar to academic papers. In this way you can get citations.
not only for your papers but also for the underlying outputs. If you are an early career researcher, it may take some time to build up a nice publication list. By publishing all your outputs, you can not only share the full breadth of your research (not just the snapshot presented in the research article) but also get more credit for your hard work and acquire increased visibility and impact.

In addition, you can also obtain a persistent identifier for yourself as a researcher. If you sign up for an ORCID iD 44 (entirely free and not dependent on your research institution), you can ensure that all your outputs are linked to you and collected in one location, regardless of changes in your surname, email address, or affiliation.

**Get a License**

Additionally, repositories offer a selection of licenses. By adding a license, you can make it clear to future users what they are allowed to do with your outputs and whether there are any restrictions on reuse. In general, Creative Common Licenses are suitable for data, while MIT, BSD, or Apache licenses are suitable for software. Choosing a license should be done with care. 45 If in doubt, you should discuss this with all stakeholders involved in your research project and ask for support from the appropriate legal department. Be aware that without a clear license, data and code are protected by basic copyright laws, which are different in each country. Therefore, reusing data or code that does not have a license may make redistribution of your work (or part of it) more difficult.

**Share Your Protocols**

Many research projects start with reproducing the outcomes of an existing study with the aim of building on top of it. However, the information provided in the Materials & Methods section of a publication is mostly limited, and statements such as “Contact the author for details” are not uncommon. As a result, many researchers spend months optimizing a protocol before they can get started with their project. You can share your detailed experimental and computational protocols on Protocols.io, which is a free and open platform. 46 As you optimize your protocol, you can add new versions and link them to your original protocol. Each published protocol gets a DOI and can be linked to existing academic publications. In this way you can obtain citations for your protocols and save others months of work.

**Share Your Analysis Scripts and Research Software**

Using a version control system is particularly crucial if you are developing analysis scripts and research software. Version control systems can be paired with online sharing platforms such as GitLab, GitHub, or GitBucket, making it even easier to collaborate, share, and archive your work. For example, there is a direct integration between GitHub and Zenodo, which is a general-purpose repository. By using the GitHub/Zenodo integration, you can get a DOI for your software and make it citable. 50 Even if you do not use version control systems, you can always upload your code to Zenodo or another repository directly through their web interface.

**OPEN ASPECTS OF THE REPRODUCIBILITY DISCUSSION**

We have provided you with multiple tips for carrying out reproducible research. However, it is worth stressing that the discussion about what exactly does it mean for research to be reproducible is ongoing. For example, researchers recently tried to reproduce 100 experiments in the field of psychological science. Researchers who participated in this study reflected that a reproduction experiment is but one experiment: “It is also too easy to conclude that a failure to reproduce a result means that the original evidence was a false positive.” 51 This large-scale reproducibility study was quickly followed by remarks and comments on the statistical approach used to assess whether a replication attempt was successful or not. For instance, some suggested that the results of the reproducibility study actually showed a fairly high reproducibility rate. Camerer et al. 52 further elaborated on the challenges of trying to reproduce reported results. The difficulty can come from testing multiple conclusions in a single paper, subtle changes in the experimental protocol, or differences in participant profiles in social science experiments. In addition, the statistical analysis approach used to estimate success or failure of the result of reproduction can also be questioned.

Therefore, the discussion about reproducibility is in a rather early phase. What are the criteria to judge whether a replication was successful? How many experiments do we need to consolidate results? “More” appears to be our current answer. The way science is executed might change in the future to accommodate for such observations. Working practices might need to be adjusted, as well as what is meant by reproducible research.

That said, you do not need to wait for this discussion to be over to start improving your working practices. There are easy ways to improve the reproducibility of your experiments.

**IT IS ALL IN YOUR HANDS**

There is a lot that can be done to improve research reproducibility, and there are two important things one should remember when thinking about it.

First, research reproducibility should be considered as a process. There are many steps that could be taken, and, as explained before, there is no single definition of “reproducible research.” So perhaps it is better to think about small steps and efforts that can be made today to make your research more reproducible. This should not be seen as a daunting task. Start small and build on as you go along.

Second, sometimes people wonder whose responsibility it should be to tackle the reproducibility crisis. Should we wait for funders to only fund reproducible research? Should we wait for the publishers to start publishing only reproducible papers? Should the institutions only hire and promote researchers who adhere to reproducible research practices? There are multiple stakeholders involved, all of whom could play a role in improving research reproducibility.

However, the key stakeholders involved, and at the same time the main beneficiaries, are the researchers themselves. So perhaps instead of thinking about whose responsibility it should be and who should take the first steps, it is better to consider the benefits of reproducible research practices. Florian Markowitz has written a useful article in Genome Biology with a suggestive title, “Five selfish reasons to work reproducibly” 54 (and for those of you who would rather watch, than read, there is also a video recording of Florian’s talk on the same topic 55). These five reasons are:
● Reason 1: reproducibility helps to avoid disaster
● Reason 2: reproducibility makes it easier to write papers
● Reason 3: reproducibility helps reviewers see it your way
● Reason 4: reproducibility enables continuity of your work
● Reason 5: reproducibility helps to build your reputation

Thus, those who invest in improving the reproducibility of their research will be the first and foremost beneficiaries of these efforts.

So don’t wait for the others to make a move, and instead think about what you can do today to improve the reproducibility of your own research practices. Every little step helps, and you will be the one who will benefit the most.

ADDITIONAL RESOURCES

Finally, if you would like to learn more or join a community or an initiative aiming to improve research reproducibility, here is a list.

● The Turing Way: A Handbook for Reproducible Data Science16
● Software and Data Carpentry lessons:29,30 collaboratively developed and freely available lessons for teaching basic lab skills for research computing and universal data literacy
● Good enough practices in scientific computing:57 article about “good enough” practices that can be implemented immediately after a Software or Data Carpentry workshop
● ReproducibilityTea:58 a grassroots journal club initiative to discuss diverse issues, papers, and ideas about improving science, reproducibility, and the Open Science movement
● ReproHack:59 a hackathon to reproduce published research using publicly available associated code and data

AUTHOR CONTRIBUTIONS

Conceptualization, N.D., M.T., and Y.T.-v.d.V.; Writing – Original Draft, N.D., M.T., and Y.T.-v.d.V.; Writing – Review & Editing, N.D., M.T., and Y.T.-v.d.V.

REFERENCES

1. Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. Nature 533, 452–454.
2. Begley, C.G., and Ioannidis, J.P. (2015). Reproducibility in science: improving the standard for basic and preclinical research. Circ. Res. 116, 116–126.
3. Shekelle, P.G., Kahan, J.P., Bernstein, S.J., Leape, L.L., Kamberb, C.J., and Park, R.E. (1998). The reproducibility of a method to identify the over-use and underuse of medical procedures. N. Engl. J. Med. 338, 1888–1895.
4. Wellcome Trust (2017). Data, software and materials management and sharing policy. https://wellcome.ac.uk/grant-funding/guidance/data-software-materials-management-and-sharing-policy%20.
5. European Commission (2016). Guidelines on FAIR data management in horizon 2020. https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/h2020-hi-oa-data-mgt_en.pdf%20.
6. Springer Nature. Research data policies. https://www.springernature.com/gp/authors/research-data-policy%20.
7. PLOS ONE (2019). PLOS ONE data availability. https://journals.plos.org/plosone/s/data-availability%20.
8. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3, 160018.
9. Vines, T.H., Albert, A.Y.K., Andrew, R.L., Débarre, F., Bock, D.G., Franklin, M.T., Gilbert, K.J., Moore, J.-S., Renault, S., and Rennison, D.J. (2014). The availability of research data declines rapidly with article age. Curr. Biol. 24, 94–97.
10. Miyakawa, T. (2020). No raw data, no science: another possible source of the reproducibility crisis. Mol. Brain 13, https://doi.org/10.1186/s13041-020-0552-2.
11. Center for Open Science. Reports, Registered Reports: peer review before results are known to align scientific values and practices.https://cos.io/my/.
12. Center for Open Science. More information on registered reports, https://www.cos.io/our-services/registered-reports%20.
13. Tutorials Point. Maven tutorial. https://www.tutorialspoint.com/maven/index.htm%20.
14. Python Packaging Authority. Tool recommendations. https://packaging.python.org/guides/tool-recommendations%20.
15. Docker. What is a Container?. https://www.docker.com/resources/what-container%20.
16. (2020). Project Jupyter. https://jupyter.org%20.
17. Verheul, I., Imming, M., Ringema, J., Mordant, A., van der Ploeg, J.-L., and Pronk, M. (2019). Data Stewardship on the Map: A Study of Tasks and Roles in Dutch Research Institutes (Zenodo). https://doi.org/10.5281/zenodo.2669150.
18. UK RSE Association, Nordic RSE Network, US RSE Association. https://rse.ac.uk/community/; https://nordic-rse.org/; https://us-rse.org%20.
19. Harvard Biomedical Data Management (2020). Electronic lab notebooks. https://datamanagement.hms.harvard.edu/electronic-lab-notebooks%20.
20. The Gurdon Institute (2020). Electronic lab notebooks - for prospective users. https://www.gurdon.cam.ac.uk/institute-life/computing/einguidance.
21. 4TU.Centre for Research Data (2017). Guidelines for creating a README file. https://researchdata.4tu.nl/fileadmin/user_upload/Documenten/Guidelines_for_creating_aREADME_file.pdf%20.
22. Kurkiewicz, D. (2017). Introduction to docstring. https://cran.r-project.org/web/packages/docstring/docstring_intro.html%20.
23. Goedder, D. (2001). PEP 257 – docstring conventions. https://www.python.org/dev/peps/pep-0257%20.
24. Oracle. Java platform, standard edition tools reference. https://docs.oracle.com/javase/8/docs/technotes/tools/windows/javaprinter.html%20.
25. RStudio. https://rstudio.com%20.
26. The Eclipse Foundation. https://www.eclipse.org%20.
27. Visual Studio Code. VS Code. https://code.visualstudio.com%20.
28. Reitz, K.. PEP 8—the style guide for Python code. https://pep8.org%20.
29. Software Carpentry. Our lessons. https://software-carpentry.org/lessons%20.
30. Data Carpentry. Lessons. https://datacarpentry.org/lessons/.
31. Git. Documentation. https://git-scm.com/doc%20.
32. Mercurial. Learning mercurial in workflows. https://www.mercurial-scm.org/guide%20.
33. Apache Subversion. Apache subversion documentation. https://subversion.apache.org/docs%20.
34. Software Carpentry. Version control with Git. https://swcarpentry.github.io/git-novice%20.
35. The Turing Way Community; Scriberia (2020). Illustrations from the Turing Way Book Dashes (Zenodo). https://doi.org/10.5281/zenodo.3695300.

36. FAIR Principles. FAIR guiding principles for scientific data and stewardship. https://www.go-fair.org/fair-principles/%20.

37. re3data.org. Registry of research data repositories. https://www.re3data.org/%20.

38. FAIRsharing. Databases. https://fairsharing.org/databases/%20.

39. Gene Expression Omnibus (GEO). https://www.ncbi.nlm.nih.gov/geo/

40. NOMAD. NOMAD repository and archive. http://nomad-repository.eu/

41. Pangaea. https://www.pangaea.de/%20.

42. Zenodo. https://zenodo.org/

43. figshare. https://figshare.com/

44. ORCID. https://orcid.org/

45. choosealicense.com. Choose an open source license. https://choosealicense.com/

46. protocols.io. https://www.protocols.io/

47. GitLab. https://about.gitlab.com/

48. GitHub. https://github.com/

49. Bitbucket. https://bitbucket.org/

50. GitHub Guides (2016). Making your code citable. https://guides.github.com/activities/citable-code/

51. Open Science Collaboration (2015). Estimating the reproducibility of psychological science. Science 349, aac4716.

52. Gilbert, D.T., King, G., Pettigrew, S., and Wilson, T.D. (2016). Comment on “Estimating the reproducibility of psychological science”. Science 351, 1037.

53. Camerer, C.F., Dreber, A., Holmzmeister, F., Ho, T.-H., Huber, J., Johannes-son, M., Kirchler, M., Nave, G., Nosek, B.A., Pfeiffer, T., et al. (2018). Evalu-
ating the replicability of social science experiments in Nature and Science between 2010 and 2015. Nat. Hum. Behav. 2, 637–644.

54. Markowetz, F. (2015). Five selfish reasons to work reproducibly. Genome Biol. 16, 274.

55. Markowetz, F. (2019). Five selfish reasons to work reproducibly. https://www.youtube.com/watch?v=yVT07Sukv9Q%

56. The Turing Community, Arnold, B., Bowler, L., Gibson, S., Herterich, P., Higman, R., Krystalli, A., Morley, A., O’Reilly, M., and Whitaker, K. (2019). The Turing Way: A Handbook for Reproducible Data Science (Zenodo).

57. Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., and Teal, T.K. (2017). Good enough practices in scientific computing. PLoS Comput. Biol. 13, e1005510.

58. ReproducibiliTea. https://reproducibilitea.org/

59. ReproHack Headquarters. https://reprohack.github.io/reprohack-hq/%

About the Authors

Dr. Nicolas Dintzner is the Data Steward of the Faculty of Technology, Policy, and Management at the Delft University of Technology, the Netherlands. He received his PhD in Software Engineering from the Delft University of Technology in 2017 for his research on configurable software evolution. His current activities include researcher support and training. He is a certified Software Carpentry instructor.

Dr. Marta Teperek is the Head of Research Data Services at TU Delft and the Director of 4TU.ResearchData. Marta did a PhD in molecular biology at the University of Cambridge. In 2015 she joined the Office of Scholarly Communication at the University of Cambridge, where she led the creation and development of the Research Data Management Facility. In 2017 Marta moved to TU Delft in the Netherlands, where she led the Data Stewardship Program. Marta serves on the Editorial Board of the Data Science Journal and Co-Chairs the Research Data Alliance Libraries for Research Data Interest Group.

Dr. Yasemin Turkyilmaz-van der Velden works as the Data Steward of the Faculty of Mechanical, Maritime and Materials Engineering in the Delft University of Technology in the Netherlands. She obtained her PhD degree from Erasmus Medical Center Rotterdam where she studied UV-induced DNA damage repair using proteomics and live cell imaging approaches. As a Data Steward, she provides customized and discipline-specific data management support to researchers at every stage of the research life cycle, with the eventual goal of achieving a cultural change in research data management in daily practice.