Open Science Toolbox

A Collection of Resources to Facilitate Open Science Practices

Lutz Heil

lheil@students.uni-mainz.de

LMU Open Science Center

April 14, 2018

[last edit: January 7, 2019]

OSC
LMU Open Science Center

Contents

Preface

The motivation to engage in a more transparent research and the actual implementation of Open Science practices into the research workflow can be two very distinct things. Work on the motivational side has been successfully done in various contexts (see section *Key Papers*). Here the focus rather lays on closing the intention-action gap. Providing an overview of handy Open Science tools and resources might facilitate the implementation of these practices and thereby foster Open Science on a practical level rather than on a theoretical.

So far, there is a vast body of helpful tools that can be used in order to foster Open Science practices. Without doubt, all of these tools add value to a more transparent research. But for reasons of clarity and to save time which would be consumed by trawling through the web, this toolbox aims at providing only a selection of links to these resources and tools. Our goal is to give a short overview on possibilities of how to enhance your Open Science practices without consuming too much of your time.

How can I use the toolbox? If you're time is too valuable to check out the toolbox entirely, a first start is to read the first chapter *Getting Started*. Though, the main body of the toolbox consists of the *Resources for Researchers*. Here you'll find Open Science resources ordered by the stage of the research cycle. If you are interested in teaching about Open Science or if you search for workshop material to educate yourself, you might want to jump to chapter 3: *Resources for Teaching*. Finally there is a small chapter called *Community* which can be helpful if you seek information on how to get involved and where to get Open Science news.

Context. This toolbox has been created during my work at the **Open Science Center (OSC)** of the Ludwig-Maximilians-Universität, Munich. This document is supposed to serve as a merger of all resources that are provided in the online toolbox (available on: http://www.osc.uni-muenchen.de/toolbox/index.html). Special thanks go to my supervisor Felix Schönbrodt who came up with the idea and made important contributions.

Contribute. The toolbox raises no claim to completeness, it should rather serve as selection of tools. If you think there is (1) an important tool or article that should be included, (2) a broken link or citation, or (3) any other issue or question, please feel free to contact me and contribute to this toolbox: lheil@students.uni-mainz.de

1 Getting Started

If you don't know where to start, here are 10 easy steps from Felix Schönbrodt's Open Science workshop, that you can take towards a more transparent and reproducible research practice. Every step will add a little more scientific value to your research.

1. **Create an account on OSF** (http://osf.io/)
2. **Upload the material for an existing study** (questionnaires, maybe reproducible analysis scripts) to an OSF project.
3. **Add an open license to all of your figures** (so that you can reuse them in later publications, blog posts, or presentations: „Figure available under a CC-BY4.0 license at osf.io/XXXX."
4. For the next project: **Change the consent forms** in a way that **open data** would be possible for that project (see https://osf.io/mgwk8/wiki/Consent%20form%20templates%20for%20open%20data/).
5. **Sign the PRO initiative** and expect openness (or a justification why not) if you review another paper (https://opennessinitiative.org/)
6. For the next data analysis: Practice to create **scripts for reproducible data analysis** (e.g., SPSS syntax, R scripts). All analytic steps that lead from raw data to the final results should be reproducible.
7. **Let a master student preregister his/her thesis**. Can be either a „local preregisteration", or a proper preregistration at OSF or at https://aspredicted.org/. See this workshop material for how to do a preregistration: https://osf.io/yd487/ https://osf.io/mx7yp/
8. **Do you own first preregistration**; enter the Prereg challenge and get 1000$: https://cos.io/prereg/
9. **Publish your first open data set**: Ensure anonymity, provide a codebook. See here for details: http://econtent.hogrefe.com/doi/pdf/10.1026/0033-3042/a000341
10. Team up with colleagues and **establish a local open science initiative** (enter your name and affiliation in this list and see other colleagues that want to engage in a local initiative)

- http://www.osc.uni-muenchen.de/news/os_flyer/index.html This info flyer gives a great overview on easy steps to take toward transparent research.
- Felix Henninger's Open Science knowledge base is a further helpful resource for quick answers to basic Open Science questions

2 Resources for Researchers

In this chapter we provide a collection of resources and tools for researchers to facilitate Open Science practices. The resources are further ordered by stage of the research cycle (*Plan your Study*, *Conduct your Study*, *Analyse your Data*, *Publish your Data, Materials, and Paper*). Also, a good first source of answers to Open Science questions is Felix Henninger's Open Science FAQ.

2.1 Plan your Study

Before diving into the data collection process, it might feel like Open Science practices such as pre-registration are consuming a lot of valuable time. But apart from the importance of these practices for credibility and transparency of your research, these practices can actually help organizing your project and saving time... especially when you don't need to search for resources.

Evaluate Research

This process should actually be part of the literature search before even planing your study. A quick way to evaluate the research that you found during your search is to use the following tools. They enable you to find discussions, evaluations, peer reviews, and replication reports on specific articles:

- https://www.altmetric.com/
- http://curatescience.org/
- https://pubpeer.com/ (also available as browser add-on, and integrated in the Altmetric's score)

Checklist for Research Workflow

Throughout the research cycle you can use Brian A. Nosek's 8-step Checklist for Research Workflow to control for an efficient open science research workflow.

Consent Forms

When data will be shared publicly particiants need to be informed about their anonymized data being shared. Here are some full german templates of consent forms for dyadic experiments and non-dyadic experiments (collected by the OSC). For english expressions that can

be employed in consent forms, see the COS Reproducible Research and Statistics Training's collection.

Power Analysis

- COS Reproducible Research and Statistics Training offers a vast body of resources and material on Power Analysis, e.g. an Introduction to Power Analysis by Courtney Soderberg and R code examples for conducting power analyses
- Advanced Power Analysis Workshop by Felix Schönbrodt

Registered Reports (RR)

The Center for Open Science (COS) gives detailed explanations and resources on Registered Reports as well as a list of journals that use Registered Reports as publishing format: https://cos.io/rr/. Links to the repsective guideline are provided.

Examplary guidelines for authors and reviewers: https://osf.io/pukzy/

Pre-Registration

- A pre-analysis checklist by David McKenzie
- A pre-analysis template provided by Alejandro Ganimian
- Pre-registration platform on the **Open Science Framework**: https://osf.io/prereg/
- OSF Guide to pre-registration

Other Registries for Pre-Registration

- https://aspredicted.org/
- AEA RCT Registry: The American Economic Association's registry for randomized controlled trials in fields of economics an other social sciences
- RIDIE Registry: The Registry for International Development Impact Evaluations (open to experimental and observational studies assessing the impact of development programs
- EGAP Registry: Evidence in Governance and Politics is providing an unsupervised stopgap function to store designs until the creation of a general registry for social science research

DataWiz

The Leibniz Institute of Psychology Information (ZDIP) offers an automated assistant tool for researchers to document and manage their data from the beginning of the research cycle

onwards. It includes assistance in creating data management plans and sharing data (internally as well as externally to the community). The tool, called **DataWiz**, is available in english and german. It is further attached to the DataWiz knowledge base which contains helpful information, guidelines, tools and resources for different steps in the research cycle

---

2.2   Conduct your Study

During data collection, using open software, the Open Science Framework, and other collaboration tools can not only enhance your workflow and collaboration in the research team but also create a foundation for data sharing and reproducibility. Here we selected a few resources for this stage.

Open Software

- PsychoPy: Open-source software as a free tool to run neuroscience or psychology experiments written in Python. Represents an alternative to Presentation™ or Inquisit™.
- PsyToolkit: Open-source software for running online or lab-based experiments (psychology).
- OpenSesame: Open-source tool for building experiments for psychology, neuroscience, and experimental economics. It provides an easy graphical interface for beginners and Python scripting for advanced users. Open Sesame can be integrated in the Open Science Framework.
- lab.js: created by Felix Henninger. A new open tool with graphical interface for building browser-based experiments.

Born Open Data

Data that is automatically uploaded to a repository (e.g. GitHub) including time stamps and automatically generated logs is called *born open*. Advantages can be full openness & transparency, data management solutions, and simplification of data sharing. For further information see:

- Rouder, J. N. (2016). The what, why, and how of born-open data. *Behavior research methods, 48*(3), 1062-1069. https://link.springer.com/article/10.3758/s13428-015-0630-z
- Rouder, J., Haaf, J. M., & Snyder, H. K. (2018, March 25). Minimizing Mistakes In Psychological Science. http://doi.org/10.17605/OSF.IO/GXCY5

Open Science Framework (OSF)

The OSF serves as data repository and collaboration tool. Researchers can structure their projects and controll access to their stored data. A version control system and a live-editing mode enhance workflow between collaborators. OSF also implements an add-on function which enables the user to sync a project with other tools and services such as Dataverse, Figshare, GitHub, Zotero. Here is an overview on OSF's feature set and its use in every step of the research cycle.

For a detailed tutorial on using the OSF to share research products, see:

- Soderberg, C. K. (2018). Using OSF to Share Data: A Step-by-Step Guide. Advances in Methods and Practices in Psychological Science. Advance online publication. DOI: https://doi.org/10.1177/2515245918757689

Other Collaboration Tools

- Figshare
- Dryad
- Zenodo
- GitHub
- Overleaf
- PaperHive

Data Anonymization

- The OpenAIRE2020 project developed a web-based anonymization application that helps ensuring participants' privacy rights. A respective online tool is in beta version at this point: https://amnesia.openaire.eu/index.html
- Alternative: ARX Data Anonymization Tool
- See Ruben Arslan's workshop on Maintaining privacy with open data

---

2.3   Analyse your Data

For your research to be transparent and reproducible, a vital part is to provide data cleaning instructions and analysis code (ideally by using nonproprietary software). See the PRO Initiative's basic guidelines for making your analyses public for a few information. This collection of resources aims at providing some helpful links to facilitate and improving your analysis sharing practices.

R

Analysing your data with free software such as R enhances reproducibility without the limitations of proprietary software. A common way to use R is by writing analysis code and graphics code in the RStudio interface. It further implements R Markdown, with can be used to create documents, reports, and presentations that are fully reproducible.

Here are some helpful links for researchers who want to learn R:

- RStudio links to different online guides to learning R: https://www.rstudio.com/online-learning/
- Swirl is an interactive learning tool which can be directly embedded in RStudio.
- RStudio cheat sheets: https://www.rstudio.com/resources/cheatsheets/
- Grolemund & Wickham's *R for Data Science*
- Hadley Wickham's *Tidy Data*

Reproducible Research

The *PRO Initiative* gives a few basic guidelines for authors on how to facilitate reproducibility when sharing your analyses: https://opennessinitiative.org/making-your-analyses-public/

More specific collections of useful advice on this topic can be found in the following sources:

Code

- Code and Data for the Social Sciences: A Practitioner's Guide by Matthew Gentzkow & Jesse M. Shapiro
- R for Reproducible Scientific Analysis: A course by Thomas Wright and Naupaka Zimmerman
- The Reseach Cycle course on principles on reproducible research and practical training in statistical programming with R, taught by Dale Barr and Lisa DeBruine
- If you want to make your code citable, see this GitHub Guide: https://guides.github.com/activities/citable-code/

Metadata

- The DRESS Protocall by Project TIER describes what the final documentation of your study should consist of (for empirical social sciences).
- Codebook Cookbook by Ruben Arslan: R-package (and online-tool) to create a codebook for your dataset.
- Creating a codebook within SPSS: https://libguides.library.kent.edu/SPSS/Codebooks

"Works on my Machine" Error

In his paper (available here), Nicholas Eubank makes the case for increasing reproducibility by testing files on a different computer. Testing or even sharing code via cloud-based platforms prevents deficits in reproducibility that occur when code runs on the researcher's local platform but not on others'. Avoid the so called *WOMME* by using tools like:

- RStudio Cloud (https://rstudio.cloud/)
- Code Ocean (https://codeocean.com/)

Workflow Documentation

- Rouder, Haaf, and Snyder (2018) wrote a helpful tutorial on how to organize a lab in the face of Open Science practices: https://psyarxiv.com/gxcy5
- R Markdown: Create fully reproducible documents that combine code execution and documentation. A big advantage is the variety of output format: documents (e.g. Word, PDF, interactive R notebook, HTML), presentation slides, shiny apps, websites and more. Multiple languages including R, python, and SQL can be used.
- GitHub serves as data repository and active research workflow tool. Tracking of contributions of others enables version control on your files. This tool is especially useful for a research team that collaborates on developing code.
- As one of multiple features, Open Science Framework's version control system and its live-editing mode facilitate collaboration within the research team.

p-Hacking

- *p-Hack like a pro* by Felix Schönbrodt
- *Do's and Don'ts of Data Analysis* by Felix Schönbrodt
- An overview on **questionable research practices** (QRP) by Ulrich Schimmack can be found here.
- Check out: http://shinyapps.org/apps/p-hacker/ for an interactive tool on p-hacking

Jupyter

The Jupyter Notebook is an open source web-application for interactive computing. Virtual notebooks support over 40 programming languages, can be shared, collaboratively edited and can return interactive output. Uses vary from data cleaning, transformation and visualization to machine learning, statistical modeling and more.

A tutorial on the Jupyter Notebook by Karlijn Willems is a first place-to-go for trying out the Jupyter Notebook. Also, help can be seeked at the Jupyter community (e.g. on GitHub or StackOverflow)

2.4   Publish your Data, Material, and Paper

In the publication stage of the research cycle, things can get a little tricky when you want to embrace Open Science values. The following links will help you find Open Access journals, check your publisher's policies, and guide you through licencing issues. Also, resources on sharing of data and materials are given.

FAIR Principles

Before Sharing your data, your material, and your paper, we recommend to check out the FAIR Data Principles, that aim to make your data:

- Findable
- Accessible
- Interoperable
- Re-usable

Also see our section *Reproducible Research* for further resurces on the topic of re-usabilty principles

Open Data Repositories

Global Registry of Research Data Repositories covering data repositories from different academic disciplines. You can directly search for repositories or browse the registry by several filters.

Examples of common general-purpose repositories:

- Open Science Framework
- Dryad Digital Repository
- Figshare
- Harvard Dataverse Network
- Zenodo

A list of general-purpose and domain-specific data repositories is also provided by Masuzzo P, Martens L. (2017) Do you speak open science? Resources and tips to learn the language. *PeerJ Preprints* 5:e2689v1 https://doi.org/10.7287/peerj.preprints.2689v1

How To Find Open Access Journals

- The Directory of Open Access Journals serves as directory for researchers who are looking for high-quality, peer-reviewed Open Access journals and respective articles. You will also find useful information on publication charges and licencing policies as well as links to editorial information.
- Use the CoFactor Journal Seleceting Tool to filter for journals fitting your search criteria (e.g. Open Access).
- The Eigenfactor Project provides a list of no-fee Open Access journals for all fields: http://www.eigenfactor.org/openaccess/fullfree.php

Publisher's Policies?

Sherpa/RoMEO is a database containing publishers' policies on self-archiving of journal articles. Since policies vary between publishers this tool will shed some light if you are unsure about publisher's policies.

Share Preprints

Although depending on publisher's policies, it is often possible to publish preprints of your paper. A common preprint repository is Cornell University's ArXiv which serves in many research fields.

For specific domains, here are some examples of preprint service (mostly powered by the Open Science Framework Preprints):

- Psychological Sciences: PsyArXiv
- Social Sciences: SocARXIV
- Engineering: engrXiV
- Agriculture and Allied Sciences: AgriXiv

Licencing

The *Peer Reviewers' Openness Initiative (PRO)* gives a very helpful overview over licensing issues that are prevalent when publishing your paper, data, and materials: https://opennessinitiative.org/licensing-issues/:

> "When thinking about what one can, must, or should do with Open Material and Open Data, one has to differentiate on the one hand legally-enforceable rules (which are handled with legal licenses) and, on the other hand, rules and standards of the scientific community. If a piece of work is in the public domain (e.g., a CC0 license) there is no legal requirement to give attribution, but as a scientist one still has the ethical obligation to give a proper citation. Laws,

ethics, and professional courtesy are all ways that the community can protect those that open their data."

Also, the OpenDefinition project lists licenses for content and data.

3 Resources for Teaching

If you are interested in teaching Open Science practices, see this selection of curriculum material, complete examples of online courses, and further tools to implement in an Open Science workshop.

Open Science Workshop Material of the LMU Open Science Center

Angelika Stefan, Felix Schönbrodt, and Lena Schiestel created materials for workshops on Open Science related topics:

- Open Science Introduction
- Preregistration
- Power Analysis
- Open Data / Open Materials / Privacy
- Open Access

You can use or remix the material under a CC-BY licence for your teaching purposes.

Open and Reproducible Research Trainer Space

The COS provides an Open and Reproducible Research Trainer Space which contains a vast body of material for Open Science workshops including sections on workshop planning (pre workshop checklists, flyers, email tempates, etc.) and curriculum (slides, full manual).

Open Science Training Handbook

Theoretical Open Science basics, organizational aspects, examples, practical guidance, and a further resources for each topic are provided by the Open Science Training Handbook. This is a great source for Open Science Trainer (and those who are planing on becoming one).

Examples of Workshops/MOOCs

- ***Making an Impact with Open Science by Dr. Michiel de Jong***

This course focuses on the benefits researchers can reap from following Open Science principles. After giving theoretical introductions to basic concepts and advantages of Open Science, the course also aims at giving practical training concerning data sharing, open access publishing and the use of social media. It is organized in four modules containing both video lectures and readings. Study load is around 10 hrs.

https://ocw.tudelft.nl/courses/making-impact-open-science/

- ***Open Science MOOC***

  "This MOOC is designed to help equip students and researchers with the skills they need to excel in a modern research environment. It brings together the efforts and resources of hundreds of researchers and practitioners who have all dedicated their time and experience to create a community platform to help propel research forward. The content of this MOOC is distilled into 10 core modules. Each module will comprise a complete range of resources including videos, research articles, dummy datasets and code, as well as 'homework' tasks to complete as individuals. Because you don't learn how to do Open Science by reading; you learn by doing it."

https://opensciencemooc.eu/

This course also offers a wide module-related collection of resources.

- ***Transparent and Open Social Science Research***

5 week (à 4 hours/week) full course by Berkeley Initiative for Transparency in the Social Sciences (BITSS) on the FutureLearn platform.

https://www.futurelearn.com/courses/open-social-science-research

- ***Teaching Reproducible Data Analysis in R***

In March 2017, the Institute of Neuroscience and Psychology and the School of Psychology at the University of Glasgow hosted a one-day workshop on teaching reproducible data analysis using R. The workshop curriculum and slides are openly available. Also, a further resource page links to additional material.

https://gupsych.github.io/trdair_workshop/index.html

- ***BITSS Research Transparency and Reproducibility Training (RT2)***

All public resources of this workshop holding place in Amsterdam in April 2018 are provided. It aims at giving an overview of tools and best practices in transparent and reproducible social sciences.

https://osf.io/fw28g/

Interactive Tools

http://shinyapps.org/apps/p-hacker/

4 Key Papers

This chapter should serve as a recommendation list of Open Science articles, guidelines, and other toolboxes offering helpful resources.

Transparency Guide

Klein, O., Hardwicke, T. E., Aust, F., Breuer, J., Danielsson, H.,… Frank, M. C. (2018, March 25). A practical guide for transparency in psychological science. Retrieved from psyarxiv.com/rtygm

Open Science: What, Why, and How

Spellman, B., Gilbert, E. A., & Corker, K. S. (2017, September 20). Open Science: What, Why, and How. Retrieved from psyarxiv.com/ak6jr

PRO Initiatives Guidelines

"These pages provide guidance in implementing open practices and links to resources on the web that help authors contribute to the scientific literature in a more open way." https://opennessinitiative.org/guidelines-for-authors/

Open Science FAQ

Felix Henninger's FAQ to Open Science topics offers a great resource for quick answers concerning Open Science topcis

Managing and Sharing Data: Best Practice for Researchers

UK Data Archive's guide for managing data for sharing and for ethics, consent, and licensing issues.

DataWiz Knowledge Base

Best practice guidelines, tools and resources for researchers before, during and after data collection: https://datawizkb.leibniz-psychology.org/

Resource Library by BITSS

Berkeley Initiative for Transparency in the Social Sciences: https://www.bitss.org/resource-tag/education/

Open Science MOOC Resource Library

Great collection of resources as part of an Open Science MOOC https://opensciencemooc.eu/open-science-resources/

Further Recommended Readings

- Reproducibility Project: Psychology: https://osf.io/ezcuj/
- The Replication Network: https://replicationnetwork.com

- Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine, 2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science, 22*(11), 1359-1366. Available at: http://journals.sagepub.com/doi/pdf/10.1177/0956797611417632
- Simmons, Joseph P. and Nelson, Leif D. and Simonsohn, Uri, A 21 Word Solution (October 14, 2012). Available at SSRN: https://ssrn.com/abstract=2160588 or http://dx.doi.org/10.2139/ssrn.2160588
- Masuzzo P, Martens L. (2017) Do you speak open science? Resources and tips to learn the language. *PeerJ Preprints* 5:e2689v1 https://doi.org/10.7287/peerj.preprints.2689v1
- Garret Christensen's Manual of Best Practices in Transparent Social Science

5 Community

Other Open Science Initiatives

Information on other Open Science initiatives, resources of the Open Science community and how to stay tuned on Open Science news.

- Center for Open Science
- Netzwerk der Open-Science-Initiativen (NOSI)
- List of Open Science Initiatives
- Berkeley Initiative for Transparency in the Social Sciences
- Commitment to Research Transparency and Open Science
- QUEST Center for Transforming Biomedical Research
- Open Science Initiative in NeuroScience at McGill University, Canada

French national action plan for buiding an Open Science ecosystem: https://www.etalab.gouv.fr/wp-content/uploads/2018/04/PlanOGP-FR-2018-2020-VF-FR.pdf (pp.57 / Engagement 18)

Social Media

Twitter:

- @OpenScience

Facebook:

- PsychMAP
- PsychMAD

Blogs:

- http://www.the100.ci/
- http://sometimesimwrong.typepad.com/
- Center for Open Science Blog

Other formats:

- http://www.openscienceradio.org/
- https://ask-open-science.org/