Reviews by Luke

Table of contents

Open and reproducible neuroimaging: From study inception to publication

- File: data/review/fulltext/oa-id-W4295290221.pdf
- DOI: https://doi.org/10.1016/j.neuroimage.2022.119623
- OpenAlex ID: https://openalex.org/W4295290221

General themes

- Strategies
- Overview
- Resources
- Tools

Type of paper

- Guide
- Review

Other notes

- They have a website version of their paper, hosted on GitHub, which is fantastic to see!
- The paper doesn't really get into collaboration itself that much, just tools that are open and that by default might make it easier to collaborate with others.
- Other terms included: "robustness" when same data but different methods give similar results, "generalizable" when new data and new methods give similar results when new data and new methods give similar results.
- Reviewed survey says that training and education were major barriers to adopting open science practices because of lack of knowledge and skill on the tools and processes. And because there are so many new tools, it is hard even for experienced researchers to navigate everything.
- The authors have a good strategy for how and why they list the tools and resources that they do in the review.
- BIDS for open neuroimaging data formatting.
- Hierarchical Events Descriptor (HED) for metadata standard.
- DataLad for managing data versioning.

Ten simple rules for helping newcomers become contributors to open projects

- File: data/review/fulltext/oa-id-W2972809368.pdf
- DOI: https://doi.org/10.1371/journal.pcbi.1007296
- OpenAlex ID: https://openalex.org/W2972809368

General themes

- Onboarding
- Documentation
- Contributing
- Centralization
- Knowledge management
- Communication
- Guidelines

Type of paper

- Guide

Other notes

This paper seems more targeted to general open projects and not strictly about open collaboration within research. However, it is useful information contained within on some general steps to take to encourage and continue collaboration in open projects and in as open ways as possible.

The rules that are more relevant to open collaboration are:

1. Be welcoming: Example, recognize when newcomers join and have an onboarding process or document.
2. Make governance explicit: Example, how are decisions made?
3. Keep knowledge up to date and findable: Example, where are the contributing guidelines?
4. Have a code of conduct that you enforce: Example, what are the guidelines for communication and what happens if someone breaks them?
5. Make it easy for contributors to get started: Example, like the being welcome, have an onboarding document and/or roadmap for the projects work.
6. Acknowledge all contributions: Any contribution can be meaningful if it helps further the goal of the project.

Collaborative open science as a way to reproducibility and new insights in primate cognition research

- File: data/review/fulltext/oa-id-W4251805646.pdf
- DOI: https://doi.org/10.31234/osf.io/8w7zd
- OpenAlex ID: https://openalex.org/W4251805646

General themes

- Systematic review
- Large-scale collaboration
- Large-scale infrastructure

Type of paper

- Review
- Example workflow

Other notes

- This paper wasn't the most relevant, as it didn't go into as much detail as we might like, even though it was describing exactly what we are looking for.
- Research needs a certain sample size to be able to answer certain questions. But primate research is difficult to get enough sample size for any given project. This need drove the push for more openness and sharing.
- Centralized platform for primate researchers and for sharing data.
- Use GitHub to share things.
- Projects should preregister the plans, and all data, code, and materials are uploaded to GitHub.
- Tools:
  - GitHub
  - Google Docs
  - Slack
  - OSF
- Use their website to disseminate news and findings
- Limitations:
  - Logistical. Impossible to use the exact same methods across all studies.
  - Also technical and knowledge, not enough funding to hire relevant expertise and technical resources.

Promoting FAIR Data Through Community-driven Agile Design: the Open Data Commons for Spinal Cord Injury (odc-sci.org)

- File: data/review/fulltext/oa-id-W3188722327.pdf
- DOI: https://doi.org/10.1007/s12021-021-09533-8
- OpenAlex ID: https://openalex.org/W3188722327

General themes

- Data infrastructure
- FAIR data
- Case study
- Agile design
- Templates
- Examples
- Strategies
- Software development

Type of paper

- Case study
- Example workflow

Other notes

- While the project's output is a data sharing platform, the process to get there required a high level of collaboration.
- Seems to use GitHub to at least build the website.
- Building up an open collaboration project takes years of continued work and effort.
- Took a multi-staged approach to establishing the group/community.

    - At each stage, it was slowly about introducing the concepts to a broader and broader audience and getting feedback all along the way.
- Development follows principles of agile software development by getting requirements from users, designing and developing those requirements, seeking feedback from users on what was developed, and testing the developed features.
- Use of the iterative development helped them identify and respond to issues that came up. This is something that could be something that non-software projects could use to help with collaboration and maintaining momentum and motivation.

UKRN Open Research Training Resources and Priorities Working Paper

- File: data/review/fulltext/oa-id-W4372403418.pdf
- DOI: https://doi.org/10.31219/osf.io/s2f6k
- OpenAlex ID: https://openalex.org/W4372403418

General themes

- Training
- Resources
- Tools
- Survey

Type of paper

- Survey

Other notes

- Not a super relevant paper for our review.
- Related to training, most training material are not created with open licenses, which makes it difficult to share and spread and be re-used.
- Training is an area that needs improvement, not enough of it or not relevant enough for current situation.

Eleven Strategies for Making Reproducible Research and Open Science Training the Norm at Research Institutions

- File: data/review/fulltext/oa-id-W4378611187.pdf
- DOI: https://doi.org/10.31219/osf.io/kcvra
- OpenAlex ID: https://openalex.org/W4378611187

General themes

- Strategies
- Best practices
- Review
- Community-building
- Training
- Assessment

Type of paper

- Guide
- Review

Other notes

- Useful guidelines for basic strategies to build up open science and reproducibility, but not too much on collaboration specifically.
- Strategies for training:
    - have more courses on reproducibility and open science
    - integrate skills in reproducibility and open science into existing courses
    - have more hands-on courses
    - embed a course training within a research group (e.g. like an intervention)
    - reproduce a study as part of a course group project, to also train collaborative team science
    - require curriculum to include reproducibility and open science training

Collaboration and Open Science Initiatives in Primate Research

- File: data/review/fulltext/oa-id-W3183365357.pdf
- DOI: https://doi.org/10.31219/osf.io/7c93a
- OpenAlex ID: https://openalex.org/W3183365357

General themes

- Large-scale collaboration
- Data aggregation
- Replications
- Limited resources
- Strategies
- Case study
- Communities of practice

Type of paper

- Example workflow
- Community building
- Case study

Other notes

- Mostly an example of what they did and why

- Faced with limited resources and small sample sizes, but wanting/needing more to make results generalizable, they put together this collaboration project.
- Outlines benefits of this collaboration, such as reducing the efforts of individual researchers by distributing workload, resources, and knowledge sharing/creation/management.
- Challenges to this collaboration are less papers, lower ability to cover more topics, less freedom of control and direction at the individual researcher level (since they are part of a larger group and need to have a consensus on a topic or standards of practice), higher level of rules and processes for dealing with researchers who try to take advantage of the community for their own benefit, as well as how to decide on authorship/contributorship for scientific publications.
  - A big challenge to these collaborations is sources of funding. Who pays for whose salary? How are funds distributed?
- Argues that any large-scale collaboration needs to use open practices by default.
- Something to consider when building large-scale collaboration is its continuation. If its started by one or a few people and with dedicated funding for it, what happens if that person or people leave or move on to other things? Or when the funding ends? These need to be considered at the start.
- The ManyPrimates use GitHub, at least to host their website with some general documentation and guidelines, as well as some analysis code and data for some papers they've published.
- The ManyPrimates project also set some basic standards for project file and folder structure, for pre-registering the study, etc.

Open and reproducible practices in developmental psychology research: The workflow of the WomCogDev lab as an example

- File: data/review/fulltext/oa-id-W4229452124.pdf
- DOI: https://doi.org/10.31234/osf.io/73bwu
- OpenAlex ID: https://openalex.org/W4229452124

General themes

- Example
- Workflows
- Barriers
- Lessons learned
- Tools

Type of paper

- Example of a workflow
- Guide

Other notes

- This was a very good paper and very relevant and detailed.
- More inclusive term might be "open scholarship"
- Applying openness can be very challenging when it involves data on children, but also that data on children is difficult to collect and benefits substantially from researchers using open practices, especially open data.
- The WomCogDev lab is a fairly small group (~13 people), almost all of whom are students (e.g. Masters, undergraduates, PhD students).
- General steps in their workflow:
  1. Design project
  2. Pre-register project or Registered Report
  3. Create experiment
  4. Recruit participants and collect data
  5. Analyze data
  6. Disseminate results
- Focus on 1, 4, and 5 since they say those are less discussed in the literature.
- For 1, involves writing something up by the lead researcher, getting feedback from all lab, writing detailed plan, upload to OSF, decide on authorship.
  - Making hypothesis and thinking of what all possible outcomes for results and what that means in the context of the hypothesis
  - A Data Management Plan is also created at this stage that includes details about how to be FAIR
- For 2, they start off with trying to do a Registered Report, but if it gets rejected, they will take the reviewer comments, revise and make edits, and then upload to OSF as a pre-registered protocol.
- For 3, they do not have a common programming language for creating experiments but do prioritize open source over proprietary.
- For 4, something they do to help with reproducibility is to use "data partners", which is some-

one who is in the lab but not directly involved in the specific project. Their task is to re-analyse the data to see if they come to the same conclusions as the lead author. Generally, the data partner is a PhD student or more senior.

- Tools used:
  - Slack for communication
  - Trello board
  - OSF, which has components made for literature, documents, methods, raw data, experiment details, analysis and results, manuscript, and communication (e.g. for outreach). Made private at first.
  - R or JASP for analysis
  - bioRxiv/PsyArXiv for preprints
- Philosophy:
  - More work done upfront, the less work done later.
  - Some openness is better than no openness.
  - Openness is on a spectrum.
  - Always improving and can't always do everything at once.
- Other workflows notes:
  - Each weekly lab meeting, people start with "an interesting error I found…" to encourage self-correction and open and honest dialogue and discussions
  - They have a lab manual which describes these processes.
  - Not all projects follow this workflow, nor are they open.
  - Try to build the values into the work culture
- Advantages of workflow:
  - Uncertainty is reduced
  - Reproducibility is inherent to the design and practice of the workflow
- Disadvantages of workflow:
  - Can sometimes be quite time-consuming
  - Can be difficult to get everyone on board and to follow it fairly closely
  - Steep learning curve
  - Overreliance on OSF, which has a cap on amount of data to put into it

Accelerating addiction research via Open Science and Team Science

- File: data/review/fulltext/oa-id-W4383376256.pdf
- DOI: https://doi.org/10.31234/osf.io/pbkrx
- OpenAlex ID: https://openalex.org/W4383376256

General themes

- Review
- Tools
- Best practices

Type of paper

- Review

Other notes

- Is mostly a review of best practices.
- The content that does get into collaboration is higher level and more general, for instance, "collaboration by sending us the data" and "we sent out polls to the group to make decisions".
- Does get into some limitations of doing larger-scale collaborations:
  - Needs leadership to have progress, so people need to agree on who and on how to limit abuse of power and to make decisions based on their own preferences or goals without getting input from the others
  - More people means more differences in opinions means more potential for conflict.

Ten strategies to foster open science in psychology and beyond

- File: data/review/fulltext/oa-id-W4281886503.pdf
- DOI: https://doi.org/10.31234/osf.io/c38a2
- OpenAlex ID: https://openalex.org/W4281886503

General themes

- Tips
- Guidelines
- Review
- Strategies
- Overview

Type of paper

- Community-building
- Tips
- Guide

Other notes

- Relevant paper as an overview
- Open science resources tend to be more theoretical then practical ("do this" but not "this is how you do this")
- Strategy: Start small and slowly incorporate open science practices.
  - For instance, start with uploading a preprint, then pre-register, then share notebooks/methods, then analysis code, etc.
- Strategy: Become familiar with national/institutional open science policies
- Strategy: Analyze and share successful cases of implementing open science practices
  - By sharing, you bring awareness to others about it, which can over time pressure and push for institutional/organizational change.
- Strategy: Include open science practices in your teaching
  - Getting students to be more familiar with and comfort on open science practices, in course settings and assignments, will mean they will be more comfortable applying it to real projects.
- Strategy: Embrace and use open educational resources
  - Use them, create them, publish them
- Strategy: Collaborate with others using open tools
  - Example: Write an article using R Markdown or Quarto. Use open formats for data. Use GitHub to share and collaborate on projects.
- Strategy: Develop networks of open collaboration
  - Build networks of working with others in an open way.
  - Develop standards, frameworks, and templates to follow within this network.
- Strategy: Voice your opinions
  - Write things in support of these initiatives
  - Outline values and principles to follow
- Strategy: Rethink and promote changes in assessment of scholarship
  - Push for and advocate for alternatives to assessments outside of the typical impact factor and h-index
- Strategy: Create opportunities for people to specialize in open science
- Tools:
  - OSF, GitHub, Zenodo, R, Quarto, R Markdown, Jupyter Notebooks

Open and collaborative tools for disaster management and risk reduction

- File: data/review/fulltext/oa-id-W4303191041.pdf
- DOI: https://doi.org/10.36335/vnjhm.2022(12).33-38
- OpenAlex ID: https://openalex.org/W4303191041

General themes

- Review
- Tools

Type of paper

- Review

Other notes

- This was a very short paper, without anything of relevance to our review.

UKRN ORCC Primer on Working in Open Research

- File: data/review/fulltext/oa-id-W4386723394.pdf
- DOI: https://doi.org/10.31219/osf.io/346hr
- OpenAlex ID: https://openalex.org/W4386723394

General themes

- Primer
- Review

Type of paper

- Review

Other notes

- This paper is very short and has almost nothing about doing open research and specifically nothing about open collaboration.

Lessons Learned: A Neuroimaging Research Center's Transition to Open and Reproducible Science

- File: data/review/fulltext/oa-id-W4283836446.pdf
- DOI: https://doi.org/10.31219/osf.io/fe74t

- OpenAlex ID: https://openalex.org/W4283836446

General themes

- Best practices
- Lessons learned
- Review

Type of paper

- Example of workflow
- Guide

Other notes

- Though a bit general, its a good overview of the steps they took to get more collaboration going and moving into more openness
- Best practices: Pre-registration, FAIR data, reproducible imaging, open science of workflows, methods, analysis code
- Strategies:
  - Include a long review period to identify best practices and tools before implementing any changes
  - Implement version control using Git and GitHub, so that when students move on, the knowledge isn't lost
  - Transition from Matlab to Python (or R)
  - For new projects, start from scratch using best practices
  - For older projects, update to best practices as much as possible given time and effort
  - All finished papers need to be uploaded to preprint archives
  - Convert data to BIDS format (very difficult procedure)
  - Converted in-house pipeline to standard fMRIPrep, which is a standard imaging pipeline
  - Upload data to OpenNeuro
  - Constructed a common data dictionary for data collected (very difficult)
- Tools:
  - Git, GitHub, Python
  - psyRxiv, bioRxiv, arXiv
  - BIDS data format as well as BIDS-Apps
  - fMRIPrep and Singularity
  - Deposit data on OpenNeuro
- Benefits:
  - Uploading to open records/archives shows institutional work and knowledge put into the work
  - Reduced time to completion of many tasks
  - Reduced costs
  - Minimal loss of knowledge when students leave, so knowledge can more easily be built up on more
  - Reducing errors from processing, analyzing, and data collection
  - Increased cross-group collaboration and knowledge sharing
  - Uploading preprints, using open source programming languages, depositing data to public repos, and using containerized reproducible pipelines all had low cost-high benefit ratios.

Bibliography